

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model.

Run on: September 2, 2003, 11:01:37 ; Search time 16.556 Seconds
(without alignments)
2099.101 Million cell updates/sec

Title: US-09-874-162a-5

Perfect score: 3865
Sequence: 1 MAPQHGCGGGGSGSPSAGS.....KALETDSVSGVSKOSKQKRL 739

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1177	30.3	900	SU12_DROME	Q9n1g9 drosophila
2	236	6.1	631	EMF2_ARATH	O81ey4 arabidopsis
3	217	5.6	440	VRN2_ARATH	O8w5b1 arabidopsis
4	190	4.9	692	FIS2_ARATH	O9znt9 arabidopsis
5	163.5	4.2	1709	CHD1_HUMAN	O14646 homo sapien
6	158.5	4.1	3969	HRX_HUMAN	P35527 homo sapien
7	157	4.0	622	KIC1_HUMAN	P40201 mus musculu
8	153.5	3.9	1711	CHD1_MOUSE	P23023 drosophila
9	153	3.9	549	DSX_MOUSE	P04104 mus musculu
10	153	3.9	627	K2C1_MOUSE	O04948 brachioosto
11	151	3.8	1235	IFR1_BRALA	P04933 plasmodium
12	146.5	3.8	1639	TRK1_YEAST	P13685 saccharomyc
13	146.5	3.8	1639	MSR1_PLATF	P04933 plasmodium
14	146	3.8	2230	GOC4_HUMAN	Q13339 homo sapien
15	145	3.7	643	K2C1_HUMAN	P04264 homo sapien
16	145	3.7	1744	TANA_XENLA	Q01550 xenopus lae
17	144.5	3.7	1875	MLP1_YEAST	O02455 saccharomyc
18	142.5	3.7	474	SOX4_HUMAN	O06945 homo sapien
19	142.5	3.7	2038	FSH_MOUSE	P13709 drosophila
20	140	3.6	349	GRP_MOUSE	P27383 arabidopsis
21	140	3.6	569	KIC1_MOUSE	P02335 mus musculu
22	140	3.6	843	MRN_MOUSE	P49657 drosophila
23	139.5	3.6	458	KICM_HUMAN	P13466 homo sapien
24	139.5	3.6	912	UBP3_YEAST	O01477 saccharomyc
25	139.5	3.6	1085	YARA_SCHPO	O09663 schizosacch
26	139	3.6	252	GRP1_PHAVU	P10495 phaseolus v
27	139	3.6	419	KIC4_XENLA	P05781 xenopus lae
28	139	3.6	481	LORI_MOUSE	P18165 mus musculu
29	139	3.6	1453	Z373_BOVIN	O9t123 bos taurus
30	139	3.6	1833	ZEP2_HUMAN	P31529 homo sapien
31	138	3.6	1701	MSR1_PLATF	P08369 plasmodium
32	137	3.5	641	EBN1_BBY	P03311 epstein-bar
33	136	3.5	384	GRP1_PETHY	P09789 petunia hyb

34	136	3.5	1701	MSR1_PLATF	P13819 plasmodium
35	135.5	3.5	183	GRP2_ORYSA	P29834 oryza sativ
36	135.5	3.5	1142	GIN4_YEAST	O12263 saccharomyc
37	135	3.5	1953	BN11_YEAST	P41832 saccharomyc
38	135	3.5	3038	TRIO_HUMAN	O75862 homo sapien
39	134.5	3.5	440	FXGA_CHICK	O98937 gallus gall
40	134.5	3.5	629	K2C3_HUMAN	P12035 homo sapien
41	134.5	3.5	971	Y228_BORBU	O51246 borrelia bu
42	134.5	3.5	1790	USO1_YEAST	P25386 saccharomyc
43	134	3.4	465	GRP2_PHAVU	P10496 phaseolus v
44	134	3.4	593	KIC1_HUMAN	P13645 homo sapien
45	133	3.4	526	KIC2_BOVIN	P06394 bos taurus

ALIGNMENTS

RESULT 1
SU12_DROME STANDARD; PRT; 900 AA.
AC Q9NUG9; Q8R9D8; Q9W55;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Polycomb protein Su(2)12 (Suppressor 12 of zeste protein).
GN SU(2)12 OR CG8013.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), CHARACTERIZATION, AND MUTANT
RP SU(2)12-2
RX MEDLINE=21430867; PubMed=11546753;
RA Birve A., Sengupta A.K., Beuchle D., Larsson J., Kennison J.A.,
RA Rasmussen-Lestander A., Mueller J.,
RT *Su(2)12, a novel Drosophila Polycomb group gene that is conserved in
RT vertebrates and plants.*
RL Development 128:3371-3379(2001).
RN [2]
RP SEQUENCE FROM N.A.
RP STRAIN=Berkley;
RC MEDLINE=20196006; PubMed=10731132;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon G.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Neilson C.R., Mills G.L.G.,
RA Abrell J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.T., Benos P.V., Bertan B.P., Bhandal D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burdits K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davidson L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mlshina N.V., Moberly C., Morris J., Mosher A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,

RA Spter E. Spreading A.C., Stapleton M., Strong R., Sun E.,
RA Stryker R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wessman D.A., Weinstein G.M., Weissensbach J.,
RA Williams S.W., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zheng G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Zibbe R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT The genome sequence of *Drosophila melanogaster*.
RL Science 287:2185-2195(2000).
RN [3]
RN REVISIONS, AND ALTERNATIVE SPLICING.
RN STRAIN-BERLELEY:
RC MEDLINE-22426069; PubMed-12537572.
RX Maier S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell R.S.,
RX Hradecky P., Huang Y., Kamalide J.S., Milburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu I., Bernan B.P.,
RA Bettecourt B.R., Celinker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter C., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart A.W., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
RT systematic review."
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN [4]
RN SEQUENCE FROM N.A. (ISOFORM 2).
RN STRAIN-BERLELEY: TISSUE-Bmbryo.
RC MEDLINE-22426066; PubMed-12537569.
RX Stapleton M., Carlson J.W., Brokstein P., Yu C., Champagne M.,
RA George R.A., Garin H., Krommiller B., Pacleib J.M., Park S., Wan K.H.,
RA Rubin G.M., Celinker S.E.;
RT "A *Drosophila* full-length cDNA resource."
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
RN [5]
RN IDENTIFICATION IN A ESC/E(2) COMPLEX WITH E(2): CAF1 AND ESC,
RN AND METHYLTRANSFERASE ACTIVITY OF THE COMPLEX.
RP MEDLINE-22296673; PubMed-12408863.
RX Cearnlin B., Melfi R., McCabe D., Seltz V., Imhof A., Piroetta V.,
RA "Prophila enhancer of Zeste/ESC complexes have a histone H3
RT methyltransferase activity that marks chromosomal Polycomb sites."
RL Cell 111:185-196(2002).
RN [6]
RN IDENTIFICATION IN A ESC/E(2) COMPLEX WITH E(2): CAF1 AND ESC, AND
RP METHYLTRANSFERASE ACTIVITY OF THE COMPLEX.
RX MEDLINE-22296674; PubMed-12408864.
RA Mueller J., Hart C.M., Francis N.J., Vargas M.L., Sengupta A.,
RA Wild B., Miller E.L., O'Connor M.B., Kingston R.E., Simon J.A.;
RT "Histone methyltransferase activity of a *Drosophila* Polycomb group
RT repressor complex."
RL Cell 111:197-208(2002).
RN [7]
RN FUNCTION: Polycomb group (Pcg) protein. Pcg proteins act by
RN forming multiprotein complexes, which are required to maintain the
RN transcriptionally repressive state of homeotic genes throughout
RN development. Pcg proteins are not required to initiate repression,
RN but to maintain it during later stages of development. They
RN probably act via the methylation of histones, rendering chromatin
RN heritably changed in its expressibility. Component of the Esc/E(2)
RN complex, which methylates Lys-9 and Lys-27 residues of histone H3.
RN Despite the presence of a zinc-finger, it does not bind directly
RN to DNA, the Esc/E(2) complex being probably recruited to DNA by
RN Pho. The Esc/E(2) complex is necessary but not sufficient to
RN recruit a functional Pcg repressive complex that represses target
RN genes, suggesting that the recruitment of the distinct PRC1
RN complex is also required to allow a subsequent repression.
RN [8]
RN SUBUNIT: Component of the Esc/E(2) complex, composed of Esc, E(2),
RN Su(2)12, Pp3, Caf1 and probably Pho. This complex is distinct
RN from the PRC1 complex, which contains many other Pcg proteins like
RN Pc, Ph, Pac, Su(2)2. The two complexes however cooperate and
RN interact together during the first 3 hours of development to
RN establish Pcg silencing.
RN [9]
RN SUBCELLULAR LOCATION: Nuclear.
RN [10]
RN ALTERNATIVE PRODUCTS:
RN Event-Alternative splicing: Named isoforms-2:
CC Name-1; Synonym-B;

```

CC      IsoId=Q9NMG9-1; Sequence=Dlsplayed;
CC      Name=2; Synonyms=A;
CC      IsoId=Q9NMG9-2; Sequence=VSP_007033, VSP_007034;
CC      Note=No experiment confirmation available;
CC      Note=Entry belongs to the VEPs (VAR2-EMF2-FIS2-SU(2)12) family.
CC      -1- SIMILARITY: Belongs to the VEPs (VAR2-EMF2-FIS2-SU(2)12) family.
CC      -1- SIMILARITY: Contains 1 C2H2-type zinc finger.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation at
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL; AF149047; AAF73149.1; -
DR      EMBL; AE003515; AAF49094.2; -
DR      EMBL; AE003515; AAN11641.1; -
DR      EMBL; AY069809; ALA39954.1; -
DR      FlyBase; FBgn0020887; Su(z)12.
DR      InterPro; IPR007087; znf_C2H2.
DR      SMART; SM00355; znf_C2H2; 1.
DR      PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
KW      Transcription regulation; Repressor; Developmental protein;
KW      Nuclear protein; Metal-binding; zinc; Zinc-finger;
KW      Alternative splicing.
FT      ZN_FING     411       434           C2H2-TYPE.
FT              527       603           VEFS-BOX.
FT      DOMAIN     527       603           ASN-RICH.
FT      DOMAIN     527       603           SER-RICH.
FT      DOMAIN     699       879           NTVLRKRYRSDSGPCTGCGHGSSGANRNKNNSHL
FT      VASNPPLIC  806       855           PATSNMMS->VEQADAREVLHSNAVDVGITIDECG
                                     GFGAVGVANNNGVASPVANNCVN (in isoform 2).
                                     /FtId-VSP_007033.
                                     /FtId-VSP_007034.
                                     Missing (in isoform 2).
FT      VARSPLIC   856       900           Lethality when homozygous.
FT      MUTAGEN    274       274           G->D: IN SU(2)12-2; INDICES LAVAL
FT      FT          274       274           LETALITY WHEN HOMOZYGUS.
SQ      SEQUENCE   900 AA; 100104 MW; 53BADB3C4DEC92E CRC64;

Query Match
Best Local Similarity 34.8%; Pred. No. 9.le-61;
Matches 281; Conservative 136; Mismatches 225; Indels 166; Gaps 26;

43 GGS CGC--GGSYSA SSSSAAAAAAGAAVLPYK-KPKMEHQADHELFLAQEFKPQTGYR 98
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db    15 GSANGIGTLGHGAPDASN---AGSVPTPAEGQVKLNGHQDELFPLAQEFKPQTGYR 70

QY    99 FLRRRLIAPFLRLITLYSHRSRTYIKRKTRKYDDMLSKYEKMGEQSLSLA-HL 157
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db    71 YLLRRHEHTNPLFLRLITSYMERMSRNKKRIISFYVSMLESI-----TOKSEASONYL 125

QY    158 QLRTTFGFHKDKDSPNSENEG-----NSVTLEVLIIVKYCHKKRKDVSPIROY 206
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db    126 HVITDSLHEKLPAIDNESGDLDQLCEAGESYSVETLLYITTKSRKRDSTIDFOL 185

QY    207 PTGRKQVPLIPDLNQTRKGNEPSLAYSSNEPEP-SNSHMVKSYSLLFRVTPRGREFNGM 265
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db    186 LSKSQQLVYNP---KDRVGEHATISIPLOTMRPMGEQHTL--YKLLPRIK-----V 231

QY    266 INGENTENIDYNEELPARKRNRREDGETPYAQMOTVEDKNNRLQLDGEEYVAOMQEBC 325
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db    232 LSPSTCN-DENAE TTP-NKRSRPN-EMPFESSELVEKSGSF-TTEGEYAMQLPLNST 286

QY    326 PI---SKRRWWEITLGGKRPP---PPEFGSGPTLOTLMTWGTBNQKSNAPLAKPLAT 379
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db    287 SIKSFSPKCTWETMPD-STYPLSLTYDYVOQOSPALKFKHLLT--SNSQLPEMTSAPELQ 342

QY    380 R-----NSESIAOEN-----KPGSVPTQTIIVAKESLTTLDTQRKEKDFPNE 421
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db    343 RYQVGHDAVAEMNTNNNNNNNNNNCSGLAKNSGGGNSIVC-----KTPP-- 386

QY    422 NRQRLRTFYOLFVNNNTROOTEARDDLHCWPACTLNCRRLYSLLNLHLKLSRFITFNYYH 481

```

```

Db      387 --EHQIYINEMTSNNTRQOTYTOBLNCPWGLDCLRLYLKHLKLCCHARNFTYOPA 444
OY      482 PKGADIVDSINCYDGSYAGNPQDIIHROGFASFERN-GPKRTPITHIVCPKRTKAM 540
Db      445 GSGADIDVTINAYGSGVAGSPYDLAAGSGSSPARGCVPRTSVSLAVCPRRKOTCL 504
OY      541 SEPLESEDEGEVQOQRTSSGHNRLYFHSSTCLPLRPOEVEVSEDEKDEPMLEKTTIOI 600
Db      505 DEFELEDEDEISNOSXYITGHNRLYHHTETCLPVHREKIDIDSEGESDEPLMLKOTIQMI 564
OY      601 BEESVNEGEKEMVLMNLMVHMKHGFADONMNMCMLEVENTGCKITIKLCRNEMML 660
Db      565 DEFSVNEGEKELMMLMNMVHMKHGFVGCQPLPLACEMFLDANGTEIVKRNLYRNLHM 624
OY      661 VSMHDFNLISIMSIDKAVTKLEMOOKLEKES----- 693
Db      625 GCLFDGLLAETVYKTVQKGLGLSKYAGGELMQROREROLKYLVDGMHKKOBDPT 684
OY      694 -----ASPANEITE-----EONSTANG 711
Db      685 LKSPKPPADPADQASTSSASTSGSGSSSQMPKRMPAHLKRGSAASSPGVQSKGTENG 744
OY      712 FSEINSKEKALETDSYSGVSKSKOKXL 739
Db      745 TNGSNS-----SSNSKNVAKRSADQL 767

```

RESULT 2

```

EMF2_ARATH STANDARD; PRT; 631 AA.
AC 08L6T4; Q93V59; Q9LU50; Q9LU51;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Polycomb protein EMBRAYONIC FLOWER 2.
GN EMF2 OR AT5G51230/AT5G51240 OR MMD22.18/MMD22.19.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC euroside II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC STRAIN=cv. Columbia;
RX Pubmed=11701882;
RA Yoshida N., Yanai Y., Chen L., Kato Y., Hiratsuka J., Miwa T.,
RT *EMBRAYONIC FLOWER2, a novel polycomb group protein homolog, mediates
RL shoot development and flowering in Arabidopsis."
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20181125; Pubmed=10718197;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RT Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC
RT clones."
RL DNA Res. 7:31-63(2000).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC STRAIN=cv. Columbia;
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "R1XN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
RT SSP consortium (Salk/Stanford/PGEC)."
RL Submitted (Aug-2002) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Polycomb group (PcG) protein. Involved in flowering
CC processes by repressing unknown target genes and preventing
CC reproductive development. PcG proteins act by forming multiprotein
CC complexes, which are required to maintain the transcriptionally
CC repressive state of homeotic genes throughout development. PcG

```

```

CC proteins are not required to initiate repression, but to maintain
CC it during later stages of development. They probably act via the
CC methylation of histones, rendering chromatin heritably changed in
CC its expressibility.
CC -1- SUBUNIT: In plants, PcG complexes are probably composed of a
CC member of the E2 family (CLF or NAR), FIE, and a member of the
CC VEFs family (FIS2, VERN2 or EMF2) (by similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms-2;
CC Name-1:
CC IsoId=Q8L6V4-1; Sequence=Displayed;
CC Name-2:
CC IsoId=Q8L6V4-2; Sequence=VSP_007456;
CC Note=No experimental confirmation available;
CC -1- TISSUE SPECIFICITY: Widely expressed throughout the life cycle
CC with higher levels in proliferating tissues. Expressed in both
CC vegetative and the reproductive shoot meristems.
CC -1- DEVELOPMENTAL STAGE: Expressed in the developing embryos and
CC endosperm, then decreases when embryos mature and soon after
CC cellularization in the shoot apical meristems (SAMs), leaf primordia,
CC and young leaves. In the reproductive shoots, it is expressed in
CC both the inflorescence and floral meristems. Later, it is
CC expressed in floral organ primordia. In coliflorescences, it is
CC expressed in SAMs and lateral organs. In roots, it is expressed in
CC root tips.
CC -1- SIMILARITY: Belongs to the VEFs (VER2-EMF2-FIS2-(2/2)12) family.
CC -1- SIMILARITY: Contains 1 C2H2-type zinc finger.
CC -1- CAUTION: Ref.2 sequences differ from that shown due to erroneous
CC gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB053171; BAB58956.1; -.
DR EMBL: AB053262; BAB58957.1; -.
DR EMBL: AB023044; BAA97386.1; ALT_SEQ.
DR EMBL: AB023044; BAA97387.1; ALT_SEQ.
DR EMBL: AY140086; AAM98227.1; -.
DR InterPro: IPR007087; Znf_C2H2.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_1; 1.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; FALSE_NEG.
KW Transcription regulation; Repressor; Flowering; Nuclear protein;
KW Zinc; Zinc-finger; Metal-binding; Alternative splicing.
FT ZN_FING 324 347
FT DOMAIN 505 583 VEFs-BOX.
FT DOMAIN 332 335 POLY-GLU.
FT DOMAIN 289 292 POLY-SER.
FT DOMAIN 396 399 POLY-ARG.
FT VARSPLIC 392 396 Missing (in isoform 2).
FT FTid=VSP_007456.
SQ SEQUENCE 631 AA; 71680 MW; F36FE92D0F62E610 CRC64;
Query Match 6.18; Score 236; DB 1; Length 631;
Best Local Similarity 19.48; Pred. No. 1,4e-06;
Matches 137; Conservative 104; Mismatches 258; Indels 208; Gaps 24;
OY 77 EHVOADHELEFLAEPKPTQYRFLRNTLAPFLHETLYMHSRNRTRIKRTRFVDD 136
Db 34 EELAAEES--LAAYCKVEELYNIIQRAINPLFLORCLRYKLEAKRRIQMTVFL--- 88
OY 137 MLKVEKMGKEQSHSIAHLQLTPTGFFHKNDKPSPSNEQNSVTLEVLVYVCCKKR 196
Db 89 -----SGAIDACVO-----TQRLPFLYTLARLVSPK- 115
OY 197 KDVSCTPIRQ-----VPIGKKQVP-----LTPDLN-----QRRPN-----FP 228

```

Db 116 -----PAEISAVYRSBACLTGGLGVGVGSAQANFLPDMNRLLLEAKSGSLAILEP 170
 QY 229 SLAVSNSEFSPNSHMYKSYSLFRVTPGRREFRNGINGETINENIDVNEELPARRRMR 288
 Db 171 SFAGANQNSQFIDSGKIHLS-----GNGIGHC-----LMSKIRLPQ----- 204
 QY 289 EDGKTYVAQMTVDKRRRLQLLDGEVAMQEMEBECPISKRAWTETILDGKRLPPEP 348
 Db 205 -----SLYASQKSPNMDLGRVDVSLVENQPCFIKXSKSEKCVS----- 247
 QY 349 FSQGPITLFTLRMTGETINDSTAPIAKPLATRNESLHOENKPGSVPTQITAYKESLTT 408
 Db 248 -IQVPSNPLT-----SSSPQOVVITISAEVSTESKPSYSSYNDIS-SSSLQ 295
 QY 409 DLQTRKEDTPNNRORLRTFYOPLYNNNTROQTAEADDLHCPCWCTLCRRLYSILRLK 468
 Db 296 IIRLR-----TGN-----VYFNRYNNKLOKTEVEDEFSQPCPLVKASFKGLRHLR 344
 QY 469 LCHSRFTFNYYHKGARIDVSTNECYDGTAGNPDIH-RQGFAP----- 514
 Db 345 STDHLLNFEEFWTEEFQAVNVSLETKETMISKV-NEDVDPRQQTFFFSKKRRRROKSQ 403
 QY 515 -----SRNGP-----VKRPITIIICRPKR----- 535
 Db 404 VRSSROGPHGLGCEVLDKTDANSVSEKSRIPGKHRYETIGAESGQVRPCTSPADV 463
 QY 536 -----TRASHSEFLESDGEVEQO--RTYSGNRLYFHSPTCLPLRPOEM-- 579
 Db 464 QSCGDPDYVOSIASSTMLQFAKTRKISIERSDLRNRSILQRFHSHRAQPMALDGLYS 523
 QY 580 EVDSEDKDPEMLLEKTIITQIEESDYNEGEKEMKLNLMHMGHGLADQNMHACMLF 639
 Db 524 DRQSEDEVDVADPEDRRMLDDVDTKDKQMHMNSVTRQRYALADSHITWACEAF 583
 QY 640 VENTGQRIK-KNLCRNFM.LHLSMHDENLISIDKAVYKLEMO 685
 Db 584 SRLGPIWVRPHILCMCRVEMVKLMNHLGLDARTMNNCNFLBOLQ 630
 RESULT 3
 VRN2_ARATH STANDARD; PRT; 440 AA.
 ID VRN2_ARATH
 AC Q8W5B1; Q23525; Q8W5B2; Q9ACF5;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Polycomb group protein VERNALIZATION 2.
 GN VRN2 OR AT4G16845 OR FCAL1.23 OR DL4450W.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_Taxid:3702;
 RN 11
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), FUNCTION, AND CHARACTERIZATION.
 RC STRAIN=cv. Landsberg erecta;
 RA MEDLINE=21575875; PubMed=11719192;
 RA Gendall A.R., Levy Y.Y., Wilson A., Dean C.;
 RT "The VERNALIZATION2 (VRN2) gene mediates the epigenetic regulation of
 RT vernalization in Arabidopsis.";.
 RL Cell 107:525-535(2001).
 RN 121
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA MEDLINE=98121113; PubMed=9461215;
 RA Bevan M., Bancroft I., Bent E., Love K., Goodman H.M., Dean C.,
 RA Bergkamp R., Dirkse W., van Staveren M., Stiekema W., Drost L.,
 RA Ridley P., Hudson S.-A., Patel K., Murphy G., Pifanelli P.,
 RA Weiler H., Weiler R., Wambolt R., Weltzienegger T., Pohl T., Terryn N.,
 RA Gleien J., Villarroel R., De Clercq R., van Montagu M., Lecharny A.,
 RA Aubourg S., Gy I., Kreis M., Lao N., Kavanagh T., Hempel S.,
 RA Kotter P., Ertlan K.-D., Rieger M., Schefer M., Funk B.,
 RA Mueller-Auer S., Silvey M., James R., Monfort A., Pons A.,

RA Puidomenech P., Douka A., Vouklatou E., Milioni D., Hatzopoulos P.,
 RA Pliavandi E., Obermaier B., Hilbert H., Duesterhoeft A., Moeres T.,
 RA Jones J.D.G., Eneva T., Palme K., Benes V., Rechmann S., Moeres W.,
 RA Cooke R., Berger C., Delseny M., Voet M., Volkert G., Menes H.-W.,
 RA Klostern S., Scheller C., Chaitwicz N.;
 RT "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
 RT Arabidopsis thaliana";.
 RL Nature 391:485-488(1998).
 RN 131
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA MEDLINE=20083488; PubMed=10617198;
 RA Mayer K.F.X., Schueller C., Wambolt R., Murphy G., Volkert G.,
 RA Pohl T., Duesterhoeft A., Stiekema W., Ertlan K.-D., Terryn N.,
 RA Harris B., Ansoerg W., Brandt P., Grivell L.A., Rieger M.,
 RA Weichselgartner M., de Simone V., Obermaier B., Maché R., Mueller M.,
 RA Kreis M., Delseny M., Puidomenech P., Watson M., Schmidheini T.,
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos P., Hoheisel J., Zimmermann W., Weiler H., Ridley P.,
 RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
 RA Van der Schueren J., Gymonprez B., Chuang Y.-J., Vandenbusche F.,
 RA Breken M., Welfjens I., Voet M., Bastiaens I., Bert R., Defoor E.,
 RA Weltzienegger T., Bothe G., Ransperger U., Hilbert H., Braun M.,
 RA Holtzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
 RA Woolman P., Klein Lankhorst R., Rose M., Haut J., Koetter P.,
 RA Beeneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
 RA De Keyser A., Buyschaert C., Gleien J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
 RA Clark L., Doggett J., Hall S., Kay M., Leonard N., Melay K., Mayes R.,
 RA Pettett A., Rajendram M.A., Lynne M., Benes V., Rechmann S.,
 RA Botkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.H.,
 RA Dose S., de Haan W., Marase A.C., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fatmann B., Grandpré K., Dauner D., Herzl A.,
 RA Neumann S., Argitlou A., Vitale D., Liguori R., Pliavandi E.,
 RA Massenet O., Quigley F., Clabaud G., Muehlen A., Felber R.,
 RA Schnabel S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
 RA Cheford F., Cooke R., Berger C., Monfort A., Casacuberta E.,
 RA Chedron T., Weber N., Vandenbol M., Baiges M., Terol J., Torres A.,
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bleike C.,
 RA Frishman D., Bevan M., Wilson R.K., de la Bastide M., Hebermann K.,
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Hebermann K.,
 RA Parnell L., Dedila N., Gnoj L., Schutz K., Huang E., Spiegel L.,
 RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threiden J.,
 RA Stoneking T., Kallick J., Graves T., Hamon G., Edwards J.,
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Krimer J., Fulton L., Mardis E., Dante M., Pein K., Hillier L.,
 RA Nelson J., Spleth J., Ryan E., Andrews S., Geisel C., Layman D.,
 RA Du H., All J., Berghoff A., Jones K., Drone K., Cotton K., Joshi C.,
 RA Antonoli B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,
 RA Ma P., Zhong Y., Preston R., Vil D., Shekhar M., Matero A., Shih R.,
 RA Svaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
 RA Granat S., Shohdy N., Hasegawa A., Hamed A., Lodhi M., Johnson A.,
 RA Chen E., Marra M., Martienssen R., McCombie W.R.;
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 RT thaliana";.
 RL Nature 402:769-777(1999).
 RN 141
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC STRAIN=cv. Columbia;
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
 RT "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
 RT SSP consortium (Salk/Stanford/PGSC).";.
 RL submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 CC -I- FUNCTION: Polycomb group (PcG) protein. Plays a central role in
 CC vernalization by maintaining repressed the homeotic gene FLC, a
 CC floral repressor, after a cold treatment. PcG proteins act by
 CC forming multiprotein complexes, which are required to maintain the
 CC transcriptionally repressive state of homeotic genes throughout
 CC development. PcG proteins are not required to initiate repression,
 CC but to maintain it during later stages of development. They
 CC probably act via the methylation of histones, rendering chromatin


```

CC heritably changed in its expressibility.
CC -1 SUBUNIT: Probable component of a PCG complex. In plants, PCG
CC complexes are probably composed of a member of the EZ family (CLF
CC or MEA), FIE, and a member of the VEFs family (FIS2, VRN2 or EME2)
CC (By similarity).
CC -1 SUBCELLULAR LOCATION: Nuclear.
CC -1 ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms-2;
CC Name=1;
CC IsoId-Q8W5B1-1; Sequence-Displayed:
CC Name=2; Synonyms-VRN2';
CC IsoId-Q8W5B1-2; Sequence-VSP_007457, VSP_007458;
CC TISSUE SPECIFICITY: Weakly expressed. Expressed both during, and
CC in the absence of vernalization.
CC -1 SIMILARITY: Belongs to the VEFs (VRN2-EME2-FIS2-SU(Z)12) family.
CC -1 SIMILARITY: Contains 1 C2H2-type zinc finger.
CC -1 CAUTION: Ref.2 and Ref.3 sequences differ from that shown due to
CC erroneous gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF284500; AAL32135.1; -
DR EMBL: AF284501; AAL32136.1; -
DR EMBL: 297342; CAB10457.1; ALT_SEQ.
DR EMBL: AL161545; CAB80935.1; ALT_SEQ.
DR EMBL: AY034902; AAK59409.1; -
DR EMBL: AY063047; AAL34221.1; -
DR InterPro: IPR007087; Znf.C2H2.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 1.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; FALSE_NEG.
KW Transcription regulation; Repressor; Nuclear protein; Zinc;
KW Zinc-finger; Metal-binding; Alternative splicing.
FT ZN_FING 86 111
FT DOMAIN 156 163 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 267 345 VEFs-BOX.
FT DOMAIN 398 402 POLY-ASN.
FT DOMAIN 411 416 POLY-ASN.
FT VARSPPLIC 102 107 GLOFHL -> VGNYYN (in isoform 2).
FT VARSPPLIC 108 440 Missing (in isoform 2).
FT VARSPPLIC 108 440 /FTId-VSP_007457.
FT VARSPPLIC 108 440 /FTId-VSP_007458.
FT VARSPPLIC 108 440 Missing (in isoform 2).
FT CONFLICT 81 81 R -> K (IN REF. 1).
FT CONFLICT 121 121 L -> S (IN REF. 1).
FT CONFLICT 185 185 A -> T (IN REF. 1).
FT CONFLICT 193 193 A -> T (IN REF. 1).
FT CONFLICT 239 239 T -> S (IN REF. 1).
FT CONFLICT 344 344 V -> A (IN REF. 1).
FT CONFLICT 355 355 C -> R (IN REF. 1).
FT CONFLICT 398 398 TSVTN -> SSDTYY (IN REF. 1).
FT CONFLICT 406 406 H -> R (IN REF. 1).
FT CONFLICT 425 425 K -> N (IN REF. 1).
FT CONFLICT 440 440 K -> KVIK (IN REF. 1).
SQ SEQUENCE 440 AA; 50623 MW; 0E0AB2C55517B66F CRC64;
Query Match 5.6%; Score 217; DB 1; Length 440;
Best Local Similarity 21.2%; Pred. No. 1.le-05;
Matches 80; Conservative 62; Mismatches 140; Indels 96; Gaps 11;

```

```

DB 173 KYCFLPDSPLANGTENCIALNDGNNRGLCYPEATELAGOPEMSTNIP---PALAHSL 229
OY 528 -----ILVCRPRKTRKASMESEFLESEGEQEQRTSSGNRRLYFSDTCLPRQEM--E 580
DB 230 DAGAKYITLTTVAVVPATKTRKLSAERSEA---RSHLLQKRFYSHRQVAPALQDWMSD 286
OY 581 VDSEDEKDEPMLREKTTIQIEEFSVDNEGEKEMKLMNLHVKHGFADNONHACMLFV 640
DB 287 ROSEDEVDVDDAFEDFROGLDDFVYVKNKEKPFMLHNSFVKORVITADGHISMAEYVS 346
OY 641 ENTGQRI-IRKNLCRNFMHLVSMDFNLISMSIDKAVTKRLREMOQLEKESAPANE 699
DB 347 RVEYKELHGYSSILFWCMRFLIKIMHGHVDSATINCNCTILENCR-----NT 394
OY 700 EITEBQNGRANGFSEINS 717
DB 395 SVTNNNNSVDHPSDSNT 412

RESULT 4
FIS2_ARATH STANDARD; PRT; 692 AA.
AC Q9ZNT9; Q9ZQPO;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE 15-SEP-2003 (Rel. 42, Last annotation update)
DE Polymorph group protein FERTILIZATION-INDEPENDENT SEED 2.
GN FIS2 OR AT2G35670 OR T20F21.14.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A., MUTANT FIS2-4, AND VARIANTS.
RC STRAIN=cv. Landsberg erecta; TISSUE=Stiliques;
RX MEDLINE=99093530; PubMed=9874812;
RA Luo M., Billoreau P., Koltunow A., Dennis E.S., Peacock W.J.,
RA Chaudhury A.;
RT "Genes controlling fertilization-independent seed development in
RT Arabidopsis thaliana.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:296-301(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA Moffat K.S., Cronin L.A., Shen M., Pal G., Van Aken S., Umayam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:761-768(1999).
RN [3]
RP SUBCELLULAR LOCATION, TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
RX PubMed=10962025;
RA Luo M., Billoreau P., Dennis E.S., Peacock W.J., Chaudhury A.;
RT "Expression and parent-of-origin effects for FIS2, MEA, and FIE in the
RT endosperm and embryo of developing Arabidopsis seeds.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:10637-10642(2000).
CC -1- FUNCTION: Polymorph group (PCG) protein. PCG proteins act by
CC forming multiprotein complexes, which are required to maintain the
CC transcriptional repressive state of homeotic genes throughout
CC development. PCG proteins are not required to initiate repression,
CC but to maintain it during later stages of development. They
CC probably act via the methylation of histones, rendering chromatin
CC heritably changed in its expressibility. Required to prevent the
CC proliferation of the central cell by repressing unknown target

```

[illegible]

OY	205	-----OVPRGKKQVPLPIPLDNLQTKGNF--PSLAAVSSN-----	236
Db	134	AHSEKISDILTTTOLAIAESSERKVPYVND---GNVSSPPRAHSAEKNESTHVNDDDD	190
OY	237	-FEPSNSHMAVSYSLF---RYTRPGRRFNGMINGETNENIDNVEE--LPARRKRR	288
Db	191	VSSPPRAHSLERKNESTHVNEDNISPPKAH-----SSAKKNSTHMDVSSPPRTASSK	245
OY	289	EDGEKTFPAQNTVDK-----NRRLQLDGEVAMQMECPISKKATWTTLIDG	340
Db	246	ETSDILTTTQPAIIEPSEPKYVRGSRKQLAKRY-----KARETOPAIRES	292
OY	341	KRLDPF-----ETFEQGP-----TLQFLRWYGE-----TNDKSNAPIAKPLA	378
Db	293	SEPKVLAHNDENVSSPPRAHSLERKASDLTTTQPAIAESSERKVPYHVDENVSSPPRAHS	352
OY	379	TRNSELHQ--ENKPGSVKPTQTIAVKESLTTTDLQTRREKDP-----NENROKLRIEY	430
Db	353	SKRKNSTRKANVDNVPSPKTRSSKKTSDILTTTQPTIAESSEPKYVRHVDNVSSPPRAH	412
OY	431	QFLVNNNRQOTEARDDLCPCWCLNCKRLKLSLKHLKLCISRFIFNVYHPKGRIDVS	490
Db	413	SSKKNKSTRKND--DNIPSPKTRSSKKTSNILTRQ-----PAIAESPEK	456
OY	491	INECYDGSYAGNP-----QDIIHQGFA-----FSRNGPVKRTPIYH	527
Db	457	VPHVNDKDVSSTPRAHSSKKNKSTHKKDONSLPKPTSSKKTSDILATTQPAKAP---	513
OY	528	ILVCPKPTKASMEFFLESEDEGEVQOQTYSSGHNRLYFHSDDCLPLRPQEM--EVSD	585
Db	514	--SEPKVTRYSRRKELAEKCEAKRERLRK---GRQFYHSQTMQPMFEQVMSNEDSEN	567
OY	586	EKDEPMLEKRTITTOEEFSDVNEGEKVMKLMNLHYMHGFLADQNMHACMLFVE----	641
Db	568	ETDDYALDISRLRLERLVGYSKEEKRYMYLMNIFVRQRYVADGHWACEEFAKLRKE	627
OY	642	-----NYGOKIIRKNLCRNFMELHLSMDFNLISIDKATKCLREMOQKLEKES	693
Db	628	EMKSSSSPDMWMBRFRIKLMNNGLICATPFHKCTILLNSNDEA-----GQPTSSA	679
OY	694	ASPAHEITE 703	
Db	680	ANANNQOSME 689	

RESULT 5		
CHD1_HUMAN	STANDARD:	PRT: 1709 AA.
ID	CHD1_HUMAN	
AC	O14646;	
DT	15-JUL-1998 (Rel. 36, Created)	
DT	15-JUL-1998 (Rel. 36, Last sequence update)	
DT	28-FEB-2003 (Rel. 41, Last annotation update)	
DE	Chromodomain-helicase-DNA-binding protein 1 (CHD-1).	
GN	CHD1.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=97470991; PubMed=9326634;	
RA	Woodage T., Basrai M.A., Baxevanis A.D., Hieter P., Collins F.S.;	
RT	"Characterization of the CHD family of proteins."	
RL	Proc. Natl. Acad. Sci. U.S.A. 94:11472-11477(1997).	
CC	-1- FUNCTION: SEQUENCE-SELECTIVE DNA-BINDING PROTEIN. COULD PLAY AN	
CC	IMPORTANT ROLE IN GENE REGULATION.	
CC	-1- SUBCELLULAR LOCATION: Nuclear.	
CC	-1- SIMILARITY: BELONGS TO THE SNP2/RAD54 HELICASE FAMILY.	
CC	-1- SIMILARITY: Contains 2 chromo domains.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -	

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch).

EMBL: AF006513; AAB87381.1; -
HSSP: P23197; IAP0
DR GeneW; HGNC:1915; CMD1.
DR M1F; 602118; -
DR GO; GO:0004003; F:ATP dependent DNA helicase activity; TAS.
DR GO; GO:0007001; P:chromosome organization and biogenesis (sen. .; TAS.
DR GO; GO:0006357; P:regulation of transcription from Pol II pro. .; TAS.
DR InterPro; IPR000953; Chromo.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR000330; SRF2_N.
DR Pfam; PF00385; chromo. 2.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00176; SRF2_N; 1.
DR SMART; SM00239; CHROMO; 2.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC_C; 1.
DR PROSITE; PS00598; CHROMO_1; 2.
DR PROSITE; PS50013; CHROMO_2; 2.
DR DNA-binding; ATP-binding; Helicase; Nuclear protein; Repeat.
KT DOMAIN 1 70 SER-RICH.
FT DOMAIN 117 137 SER-RICH.
FT DOMAIN 272 364 CHROMO 1.
FT DOMAIN 389 452 CHROMO 2.
FT NP_BIND 506 513 ATP (POTENTIAL).
FT SITE 614 617 DEAD BOX.
FT DOMAIN 1628 1644 3 x 5 AA REPEATS OF H-S-D-H-R.
FT REPEAT 1628 1632 1.
FT REPEAT 1634 1638 2.
FT REPEAT 1640 1644 3.
SQ SEQUENCE 1709 AA; 196517 MW; 416409C913D6A935 CRC64;

Query Match 4.2%; Score 163.5; DB 1; Length 1709;
Best Local Similarity 20.4%; Pred. NO. 0.074;
Matches 128; Conservative 79; Mismatches 174; Indels 245; Gaps 32;

OY 14 SGPSANGSGGGFGGSAANAATATASGGKGGSCGGGGYSASSSSSAAAAAGAVLPYK 73
DB 24 SGSASGSGSGSSSSSDGSSSSOS-----GSSDSDSGSGSGSSQSSSESDTGRNKVQAKP 78
OY 74 PKMEHYQADHLLPLQAFKPTQIYRFLTRNLIAPIFLRLTYLTHSHNSRPNIRKRTFK 133
DB 79 PKVQGA-----FMKSSPSILAV-----QRSAILRK----- 104
OY 134 VDDMLSKVEKKMGEOESHLSAHLQLPFTGFPHKNDKPSNSENGNSVTLEVLVYKCH 193
DB 105 -----QQQQQQQQQHQASSN-----SSGEEDSSSESDSDSS-----EYKR 141
OY 194 KRRDVSCTPIQVPTPKKQVPLIPDLNTRKGNFPSLAVSSNE-----F 237
DB 142 KKHND-----EDWMSGSGS-PSOGSGDSESEEREKSSCDETESDY 182
OY 238 EPSNSHWKYSILFRTYRPPRRRENG-MINGETNEDVWELPARKRRRDEGKTFV 296
DB 183 EPKNK--VKS-----RKPNRSKSKNGKILLGKKRKQDSSE-----DDEED-- 224
OY 297 AQMTVFQDKRRRLQLDGEYEVAMQEMEB----- 324
DB 225 -----YNDKDKSSRRQATVNVSYKDEDEMKDSDDLLEVCGEDVQPEEPFEETIERMD 279
OY 325 CPISKRKATWETI-----LDGKRLPEETTSQGFLLQFLRMGTGTNDKSTAPIAKPLA 378
DB 280 CRIGKKGATGATTTTLYAEDADDPNAGCEKKNKEPEIYILLIKMGWSIHNT----- 331
OY 379 TRNSLSLQENKPPGVSVPRTQTLAVKESLTTLDLQTPKEDQFN--ENRKLRIEYQFLVNN 436
DB 332 WELETELQQVNRGKK-----LNNYKKRDETRKMLKNAPEDEVY--YN- 375

```

OY      437 NTROGTLFARDLDCPMPCTCLNCKRTKSLKHLTKLS--REFENY-VHPK-----
DB      376 ---CQDELIDDLH-----KOIOIVERT-IANSNOKSAAGIPDYICKMOGLPYSECS 4223
OY      484 ---GARIDSINECYDGSYAGNPODIHROPGFASRNGPVKRTPI--THIIVCRPKRTKA 538R
DB      423 MEDGALISKKFKQACID-EX-----IIIII---:I II:
OY      539 SMSPELESEDEVEOORTYSGGHNRL 564
DB      463 ----FV-----ALKKPSTYIGGHEGL 479

RESULT 6
HRX_HUMAN
ID      HRX_HUMAN          STANDARD;             PRT;   3969 AA.
AC      Q03164; Q13743; Q13744; Q14845; Q16364; Q9UNM3;
DT      01-OCT-1993 (Rel. 27, Created)
DT      01-NOV-1995 (Rel. 32, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Zinc finger protein HRX (ALL-1) (Trithorax-like protein).
GN      MLL OR HRX OR ALL1 OR TRAI OR HTRX.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX      NCBI_TaxId=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=93046667; PubMed=1423624;
RA      Tkachuk D.C., Kohler S., Cleary M.L.;
RT      "Involvement of a homolog of Drosophila trithorax by 11q23
RL      Cell 71:691-700(1992)."
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=96290553; PubMed=8703835;
RA      Nilsson I., Loechner K., Slegler G., Grell J., Beck J.D., Fey G.H.,
RA      Marschalek R.;
RT      "Exon/intron structure of the human ALL-1 (MLL) gene involved in
RT      translocations to chromosomal region 11q23 and acute leukemias.";
RL      Br. J. Haematol. 93:966-972(1996)."
RN      [3]
RP      SEQUENCE OF 1-1909 FROM N.A.
RX      Yamanoto K., Seco M., Komatsu H., Iida S., Akao Y., Akouma S.,
RA      Kodera Y., Nakazawa S., Ariyoshi Y., Takahashi T., Ueda R.;
RT      "Two distinct portions of Lfng19/ENL at 19p13 are involved in t(11;19)
RT      leukemia.";
RL      Oncogene 8:2617-2625(1993)."
RN      [4]
RP      SEQUENCE OF 1317-2328 FROM N.A.
RC      TISSUE=Brain;
RX      MEDLINE=93265134; PubMed=1303259;
RA      Djibali M., Sellert L., Parry P., Bower M., Young B.D., Evans G.A.;
RT      "A trithorax-like gene is interrupted by chromosome 11q23
RT      translocations in acute leukemias.";
RL      Nat. Genet. 2:113-118(1992)."
RN      [5]
RP      SEQUENCE OF 1251-1538 FROM N.A.
RX      MEDLINE=94215165; PubMed=8162575;
RA      Gu Y., Alder H., Nakamura T., Schloman S.A., Prasad R., Canaanl O.,
RA      Saito H., Croce C.M., Canaanl E.;
RT      "Sequence analysis of the breakpoint cluster region in the ALL-1 gene
RT      involved in acute leukemia.";
RL      Cancer Res. 54:2326-2330(1994)."
RN      [6]
RP      SEQUENCE OF 1251-1654 FROM N.A. (ISOFORM 14P-18B).
RX      MEDLINE=95322025; PubMed=7598802;
RA      Mbongollo D., Burnett R., McCabe N., Thirman M., Gill H., Yu H.,
RA      Rowley J.D., Diaz M.O.;
RT      "The human MLL gene, nucleotide sequence, homology to the Drosophila
RT      trit-zinc-finger domain, and alternative splicing";

```

RL DNA Cell Biol. 14:475-483(1995).
 RN [7]
 RP SEQUENCE OF 1212-1603 FROM N.A.
 RX MEDLINE-95315013; PubMed-7794749;
 RA Maschiale R., Grell J., Lochner K., Nilsson I., Stegler G.,
 RA Zweckbronner I., Beck J.D., Fey G.H.;
 RT "Molecular analysis of the chromosomal breakpoint and fusion
 transcripts in the acute lymphoblastic SEM cell line with chromosomal
 translocation t(4;11)."
 RL Br. J. Haematol. 90:308-320(1995).
 RN [8]
 RP SEQUENCE OF 1421-1540 FROM N.A.
 RX MEDLINE-94020842; PubMed-8414518;
 RA Forster A., Rabbitts T.H.;
 RT "A method for identifying genes within yeast artificial chromosomes:
 application to isolation of MLL fusion cDNAs from acute leukaemia
 translocations."
 RL Oncogene 8:3157-3160(1993).
 RN [9]
 RP CHROMOSOMAL TRANSLOCATION WITH GAS7.
 RX MEDLINE-20183971; PubMed-10706619;
 RA Megonigal M.D., Cheung N.-K.V., Rappaport E.F., Nowell P.C.,
 RA Wilson R.B., Jones D.H., Addya K., Leonard D.G.B., Kushner B.H.,
 RA Williams T.M., Lange B.J., Felix C.A.;
 RT "Detection of leukemia-associated MLL-GAS7 translocation early during
 chemotherapy with DNA topoisomerase II inhibitors."
 RL Proc. Natl. Acad. Sci. U.S.A. 97:2814-2819(2000).
 RN [10]
 RP CHROMOSOMAL TRANSLOCATION WITH AF3P21.
 RX MEDLINE-20115194; PubMed-10648423;
 RA Sano K., Hayakawa A., Piao J.-H., Kosaaka Y., Nakamura H.;
 RT "Novel SH3 protein encoded by the AF3P21 gene is fused to the mixed
 lineage leukemia protein in a therapy-related leukemia with
 t(3;11)(p21;q23)."
 RL Blood 95:1066-1068(2000).
 RN [11]
 RP INTERACTION WITH SBFL.
 RX MEDLINE-98196760; PubMed-9537414;
 RA Cui X., De Vivo I., Slany R., Miyamoto A., Firestein R., Cleary M.L.;
 RT "Association of SET domain and myotubularin-related proteins modulates
 growth control."
 RL Nat. Genet. 18:331-337(1998).
 CC -1- FUNCTION: Possibly acts as a transcriptional regulatory factor.
 CC -1- SUBUNIT: Interacts with SBFL.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named Isoforms-2;
 CC Name-1;
 CC IsoId-Q03164-1; Sequence-Displayed;
 CC Name-14P-18B;
 CC IsoId-Q03164-2; Sequence-VSP_006666;
 CC -1- TISSUE SPECIFICITY: Heart, lung, brain and T and B lymphocytes.
 CC -1- DISEASE: Involved in acute leukemias by chromosomal translocations
 CC t(11;19)(q23;p13.3) that involves MLL and MLLT10/ENL;
 CC t(4;11)(q21;q23) that involves MLL and MLLT2/AF4; t(9;11)(p22;q23)
 CC that involves MLL and MLLT3/AF9; t(6;11)(q27;q23) that involves
 CC MLL and MLLT4/AF6; t(11;17)(q23;q21) that involves MLL and
 CC MLLT6/AF17; t(x;11)(q13;q23) that involves MLL and MLLT7/AFX1;
 CC t(10;11)(p13;q23) that involves MLL and MLLT10/AF10;
 CC t(11;11)(q21;q23) that involves MLL and AF10; t(11;19)(q23;p13.3)
 CC that involves MLL and ELN; t(11;19)(q23;p23) that involves MLL and
 CC GAS7; and t(3;11)(p21;q23) that involves MLL and AF3P21.
 CC -1- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
 CC -1- SIMILARITY: Contains 1 bromodomain.
 CC -1- SIMILARITY: Contains 3 PHD-type zinc fingers.
 CC -1- SIMILARITY: Contains 1 CXXC-type zinc finger.
 CC -1- SIMILARITY: Contains 1 post-SET domain.
 CC -1- SIMILARITY: Contains 1 CXXC-type zinc finger.
 CC -1- DATA BASE: NAME-Atlas Genet. Cytogenet. Oncol. Haematol.;
 CC WWW="http://www.intoblogen.fr/services/chromosome/genes/MLL.html".
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L04284; AA58669.1; -;
 DR EMBL; 269744; CA93625.1; -;
 DR EMBL; 269745; CA93625.1; JOINED.
 DR EMBL; 269746; CA93625.1; JOINED.
 DR EMBL; 269747; CA93625.1; JOINED.
 DR EMBL; 269748; CA93625.1; JOINED.
 DR EMBL; 269749; CA93625.1; JOINED.
 DR EMBL; 269750; CA93625.1; JOINED.
 DR EMBL; 269751; CA93625.1; JOINED.
 DR EMBL; 269752; CA93625.1; JOINED.
 DR EMBL; 269753; CA93625.1; JOINED.
 DR EMBL; 269754; CA93625.1; JOINED.
 DR EMBL; 269755; CA93625.1; JOINED.
 DR EMBL; 269756; CA93625.1; JOINED.
 DR EMBL; 269757; CA93625.1; JOINED.
 DR EMBL; 269758; CA93625.1; JOINED.
 DR EMBL; 269759; CA93625.1; JOINED.
 DR EMBL; 269760; CA93625.1; JOINED.
 DR EMBL; 269761; CA93625.1; JOINED.
 DR EMBL; 269762; CA93625.1; JOINED.
 DR EMBL; 269763; CA93625.1; JOINED.
 DR EMBL; 269764; CA93625.1; JOINED.
 DR EMBL; 269765; CA93625.1; JOINED.
 DR EMBL; 269766; CA93625.1; JOINED.
 DR EMBL; 269767; CA93625.1; JOINED.
 DR EMBL; 269768; CA93625.1; JOINED.
 DR EMBL; 269769; CA93625.1; JOINED.
 DR EMBL; 269770; CA93625.1; JOINED.
 DR EMBL; 269771; CA93625.1; JOINED.
 DR EMBL; 269772; CA93625.1; JOINED.
 DR EMBL; 269773; CA93625.1; JOINED.
 DR EMBL; 269774; CA93625.1; JOINED.
 DR EMBL; 269775; CA93625.1; JOINED.
 DR EMBL; 269776; CA93625.1; JOINED.
 DR EMBL; 269777; CA93625.1; JOINED.
 DR EMBL; 269778; CA93625.1; JOINED.
 DR EMBL; 269779; CA93625.1; JOINED.
 DR EMBL; 269780; CA93625.1; JOINED.
 DR EMBL; D14540; BAA03407.1; -;
 DR EMBL; L01986; AAA9251.1; -;
 DR EMBL; U04737; AAA18644.1; -;
 DR EMBL; S78570; AAB34770.1; -;
 DR EMBL; X83604; CAA58584.1; -;
 DR EMBL; S66432; AAB28545.1; -;
 DR EMBL; AF231998; AAG26332.2; ALT_TERM.
 DR PIR; A44265; A44265.
 DR TRASNFPAC; T02337; -;
 DR Genew; HGNC:7132; MLL.
 DR MIM; 159555; -;
 DR GO; GO:0005634; C:nucleus; TAS.
 DR GO; GO:0005702; F:RNA polymerase II transcription factor acti.; TAS.
 DR GO; GO:0007048; P:oncogenesis; TAS.
 DR GO; GO:0006366; P:transcription from Pol II promoter; TAS.
 DR InterPro; IPR001487; Bromodomain.
 DR InterPro; IPR003889; FYRICH_C.
 DR InterPro; IPR003888; FYRICH_N.
 DR InterPro; IPR003616; PostSET.
 DR InterPro; IPR002857; ZnF_CXXC.
 DR InterPro; IPR001965; ZnF_PHD.
 DR Pfam; PF02178; AT_hook; 2.
 DR Pfam; PF00628; PHD; 3.
 DR Pfam; PF00856; SET; 1.
 DR Pfam; PF02008; zf-CXXC; 1.
 DR SMART; SMO0297; BROMO; 1.
 Query Match 4.18; Score 158.5; DB 1; Length 3969;
 Best Local Similarity 20.48; Pred. No. 0.41;

Matches 179; Conservative 97; Mismatches 313; Indels 287; Gaps 40;

OY 3 POKHGGGGGGG-----SGPSAGSGGGGFGGS-AAVAATAGGKSGC 43
 DB 13 PGTGGGGGGGGGGLGXPGRQVPALLLPRLPPVGGGGGAPPPSPAVAAAAASSSG 72
 OY 44 GSCGGGGGTSASSSSSAAAAGANVLPVKKPKMEHQADHEFLQAEKPTQIYRLRTR 103
 DB 73 GVPGGGAAAASASSSSSSSSSSSSSSSGPALLRGPFGDALQ-----FTG 117
 OY 104 NLIAPIFLHRTLYNHSRFTNKKRKVDMLSKVKMGGEOSHLSAHLQLTFG 163
 DB 118 -VSAI-----GTLNR--FRAVFGSGGGGGGDEQ-----FLG 150
 OY 164 FPHKNDPSPNSEONSVTLVLYVCHKKRNDVSCPIRQVTKKQVPLIDNQR 223
 DB 151 F-----GSDE-----VRV-----RSPTSPSKTSPRRPRGR 178
 OY 224 P--GNFPLAVSSNE--FEPSNSHWKYSLSLFRYTRPGRRFNGMINETNIDVNEE 279
 DB 179 PMSGSDRNSAILSDPSVFSPLNKEKTS-----CDKIKKDS 215
 OY 280 LPARRKRNDEKKTVAQMTVFDKNRL--QLLDGEYVAMQEMECP-----ISK 329
 DB 216 KSEKKRGR---PPIFGYKIKITTHGKDISELPKMKEDSLKKIKRTPSFOQATKIK 272
 OY 330 KRATWETILDGKRLPPFETFSOGPTLOFTLRWTGETNDKSTAPIANPLATRNSESLHOE- 388
 DB 273 LRA-----GKLSPLKSKFTKG-KLQIGKKGVOIYRRGRGPSTRIKTPSGLLSEL 324
 OY 389 NRPGSYK-----PPTIYAKESLTDLDTRKE---KDTPENRQKLIIFYOPLYNNMT 438
 DB 325 ERQKVRKKEGTPILT---KEDKTVRQSPRRIRIKRPIIIPSKRTDARTAKOL-----Q 377
 OY 439 ROOTEARDDLHCPCWTLNCRKLYSLKHLK-----LCHSFTINYVHP- 482
 DB 378 RAKKGAKIKERKADLOGKRVKQYKNIRQITMPVYSAISSNITTPRFIDEEDIDPP 437
 OY 483 -KGARIDVINECYDSYAGNPQ-----DIHR--QPGFAPSRLNG-----518
 DB 438 IKIARLESTPNSRFSAPSCGSSSEKSSAOSHSSOMSDSSRSSPSVDSTDSQASEIQ 497
 OY 519 -----PVKRTPTIHLVCPKRTKMSSEF-----LESDGEVE 552
 DB 498 VLPEERSDTPPEVHPPLPISQSPENESNDRSRYSSEKSGRTTKKLSLOS---AP 553
 OY 553 QOETVSSGNRLYFHSDTCLPLRQMEVDSDEKDPENLRKTIQIEF-----SDVN 607
 DB 554 QOQTSSPPPLTTPP---PLOP---ASSISDHP--WLMPTIPLASFLDASTAPMQ 605
 OY 608 EGEKEVWK---LWNLHVKKHGFADONMNAHACMLEVENYGR--IIKNLCRNFMLHVS 662
 DB 606 GKXKSLRREPTFKWT--SLKHSREPOYSSA-----KAKGGLRKLPLDFNFRPPPLT 657
 OY 663 MHDENLISMSIDKAVTKLR-----EMOQ-----KLEGESASPANEITE 703
 DB 658 PEDVGFASGFSASGTAASARLFPRLHSGTRFDMHKSPILRAFRPTEASHSRIFESVTL 717
 OY 704 EONGTANGFSEINSKEKALETDSVSGYSKOKKOL 739
 DB 718 PSNRTSAGTSS-----SGVSNKRRKRV 740

RESULT 7

KICI_HUMAN STANDARD; PRT; 622 AA.

AC P35527; 014665;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Keratin, type I cytoskeletal 9 (Cytokeratin 9) (K9) (CK 9).
 GN KRT9.
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Foot sole tissue;
 RX MEDLINE=94131202; PubMed=7507869;
 RA Langbein L., Heid H.W., Moll I., Franke W.M.;
 RT "Molecular characterization of the body site-specific human epidermal
 RT cytokeratin 9: cDNA cloning, amino acid sequence, and tissue
 RT specificity of gene expression.";
 RL J. Invest. Dermatol. 103:57-72(1993).
 RN [2]
 RP SEQUENCE OF 449-465.
 RX MEDLINE=90267446; PubMed=2140676;
 RA Rosen E.M., Meromsky L., Romero R., Sotter E., Goldberg I.;
 RT "Human placenta contains an epithelial scatter protein.";
 RL Biochem. Biophys. Res. Commun. 168:1082-1088(1990).
 RN [3]
 RP VARIANTS EPPK VAL-156 AND PRO-171.
 RX MEDLINE=94274199; PubMed=7516304;
 RA Hennies H.-C., Zehender D., Kunze J., Kuester W., Reis A.;
 RT "Keratin 9 gene mutational heterogeneity in patients with
 RT epidermolytic palmoplantar keratoderma.";
 RL Hum. Genet. 93:649-654(1994).
 RN [4]
 RP VARIANT EPPK SER-160.
 RX MEDLINE=95015968; PubMed=7523529;
 RA Bonifas J.M., Matsumura K., Chen M.A., Berth-Jones J.,
 RA Hutchinson P.E., Zlotcower M., Fritsch P.O., Epstein E.H., Jr.;
 RT "Mutations of keratin 9 in two families with palmoplantar
 RT epidermolytic hyperkeratosis.";
 RL J. Invest. Dermatol. 103:474-477(1994).
 RN [5]
 RP VARIANT EPPK TYR-160.
 RX MEDLINE=94184355; PubMed=7511021;
 RA Torchard D., Blanchet-Bardon C., Serova O., Langbein L., Narod S.,
 RA Janin N., Goguel A.F., Bernheim A., Franke W.W., Lenoir G.M.,
 RA Feunteun J.;
 RT "Epidermolytic palmoplantar keratoderma cosegregates with a keratin 9
 RT mutation in a pedigree with breast and ovarian cancer.";
 RL Nat. Genet. 6:106-110(1994).
 RN [6]
 RP VARIANTS EPPK LYS-160; GLN-162 AND TRP-162.
 RX MEDLINE=94214498; PubMed=7512862;
 RA Reis A., Hennies H.-C., Langbein L., Digtved M., Mische D.,
 RA Dreschler M., Schroeck E., Royer-Pokora B., Franke W.W., Sperling K.,
 RA Kuester W.;
 RT "Keratin 9 gene mutations in epidermolytic palmoplantar keratoderma
 RT (EPPK).";
 RL Nat. Genet. 6:174-179(1994).
 RN [7]
 RP VARIANTS EPPK TRP-162 AND SER-167.
 RX MEDLINE=95164983; PubMed=7532199;
 RA Rothnagel J.A., Wojcik S., Liefer K.M., Dominey A.M., Huber M.,
 RA Hohl D., Roop D.R.;
 RT "Mutations in the 1A domain of keratin 9 in patients with
 RT epidermolytic palmoplantar keratoderma.";
 RL J. Invest. Dermatol. 104:430-433(1995).
 RN [8]
 RP VARIANT EPPK VAL-159.
 RX MEDLINE=97348961; PubMed=9204965;
 RA Endo H., Hatamochi A., Shinkai H.;
 RT "A novel mutation of a leucine residue in coil 1A of keratin 9 in
 RT epidermolytic palmoplantar keratoderma.";
 RL J. Invest. Dermatol. 109:113-115(1997).
 RN [9]
 RP VARIANTS EPPK THR-156; VAL-156 AND GLN-162.
 RX MEDLINE=99072662; PubMed=9856842;
 RA Covello S.P., Irvine A.D., McKenna K.E., Munro C.S., Nevlin N.C.,
 RA Smith F.J.D., Utto J., McLean W.H.I.;
 RT "Mutations in keratin K9 in kindreds with epidermolytic palmoplantar
 RT keratoderma and epidemiology in Northern Ireland.";

RL J. Invest. Dermatol. 111:1207-1209(1998).
 CC -1- FUNCTION: MAY SERVE AN IMPORTANT SPECIAL FUNCTION EITHER IN THE
 CC MATURE PALMAR AND PLANTAR SKIN TISSUE OR IN THE MORPHOGENIC
 CC PROGRAM OF THE FORMATION OF THESE TISSUE.
 CC -1- SUBUNIT: HEMEROPETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE TERMINALLY DIFFERENTIATED
 CC EPIDERMIS OF PALMS AND SOLES.
 CC -1- DISEASE: DEFECTS IN KRT9 ARE A CAUSE OF EPIDERMOLYTIC PALMOPANTAR
 CC KERATODERMA (EPPK), AN AUTOSOMAL DOMINANT DISEASE CHARACTERIZED BY
 CC DIFEUSE THICKENING OF THE EPIDERMIS ON THE ENTIRE SURFACE OF
 CC PALMS AND SOLES SHARPLY BORDERED WITH ERYTHEMATOUS MARGINS.
 CC -1- MISCELLANEOUS: There are two types of cytoskeletal and
 CC microfilament keratin, I (acidic) and II (neutral to basic) (40-
 CC 35 and 36-70 kDa, respectively).
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC -1- CAUTION: WAS ORIGINALLY (REF.2) THOUGHT TO BE A 60 KDA CHAIN OF
 CC PLACENTAL SCATTER PROTEIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; X75015; CAA52924.1; -;
 CC EMBL; Z29074; CAA82315.1; -;
 CC EMBL; S69510; AAC60619.1; -;
 CC PIR; I37984; I37984.
 CC Gene; HGNC:6447; KRT9.
 CC MIM; 144200; -;
 CC DR GO; GO:0005200; F:structural constituent of cytoskeleton; TAS.
 CC DR GO; GO:0008544; P:epidermal differentiation; TAS.
 CC DR InterPro; IPR001664; IF.
 CC DR InterPro; IPR002957; Keratin_I.
 CC Pfam; PF00038; filament_1.
 CC DR PRINTS; PR01248; TYPE1KERATIN.
 CC DR PROSITE; PS00226; IF; 1.
 CC DR Intermediate filament; Colled coil; Keratin; Disease mutation.
 CC KW DOMAIN 1 151 HEAD.
 CC FT DOMAIN 152 460 ROD.
 CC FT DOMAIN 461 622 TAIL.
 CC FT DOMAIN 152 187 COIL 1A.
 CC FT DOMAIN 188 206 LINKER 1.
 CC FT DOMAIN 207 298 COIL 1B.
 CC FT DOMAIN 299 321 LINKER 12.
 CC FT DOMAIN 322 460 COIL 2.
 CC FT DOMAIN 14 20 POLY-GLY.
 CC FT VARIANT 156 156 M->T (IN EPPK).
 CC FT VARIANT 156 156 /FtId-VAR_010499.
 CC FT VARIANT 159 159 M->V (IN EPPK).
 CC FT VARIANT 159 159 /FtId-VAR_010500.
 CC FT VARIANT 159 159 L->V (IN EPPK).
 CC FT VARIANT 160 160 N->Y (IN EPPK).
 CC FT VARIANT 160 160 /FtId-VAR_010501.
 CC FT VARIANT 160 160 N->K (IN EPPK).
 CC FT VARIANT 160 160 /FtId-VAR_003822.
 CC FT VARIANT 160 160 N->S (IN EPPK).
 CC FT VARIANT 160 160 /FtId-VAR_010502.
 CC FT VARIANT 160 160 N->Y (IN EPPK).
 CC FT VARIANT 160 160 /FtId-VAR_010503.
 CC FT VARIANT 162 162 R->Q (IN EPPK).
 CC FT VARIANT 162 162 /FtId-VAR_003823.
 CC FT VARIANT 162 162 /FtId-VAR_003824.
 CC FT VARIANT 162 162 R->W (IN EPPK).
 CC FT VARIANT 162 162 /FtId-VAR_003824.
 CC FT VARIANT 167 167 L->S (IN EPPK).
 CC FT VARIANT 167 167 /FtId-VAR_003825.
 CC FT VARIANT 171 171 Q->P (IN EPPK).
 CC FT VARIANT 171 171 /FtId-VAR_010504.
 CC FT CONFLICT 12 12 T->SR (IN REF. 1; CAA82315).
 CC FT SEQUENCE 622 AA; 61987 MW; 898C3825D4B5BD94 CRC64;
 CC Query Match 4.0%; Score 157; DB 1; Length 622;

Best Local Similarity 20.9%; Pred. No. 0.05;
 Matches 89; Conservative 63; Mismatches 165; Indels 108; Gaps 19;
 QY 9 GCGGGGSPMSGGGGGGGSAVAATAAGSKS-----GGGSGCG-CGSA----- 54
 DB 73 GGGGGGGSSSLGGGGGGGSRGGGGSAGGYSGGGCGGGGGGGGGGGGGGGGG 132
 QY 55 -SSSSMAAAGAAVLPVKPKMEYQADHELFL---QAFER----- 92
 DB 133 LGGGGGAGGGGGGILFLANESKSTQELNSRLASTLDVQALAEANNDLENKIQDWYDKG 192
 QY 93 PTOYR---PLRT---RLIAPFLHRTLYSHNSRNIR--KTRVDMLEK-- 140
 DB 193 PMAIKNYSPYNTIDLDKQIDVLTGNNNTLLDIDNTRLTLDFFRIKFEEMQNLOGV 252
 QY 141 -----VEKMGGEOSHLSAHLQTLTGFGFKNDKPPNSNEQNSVYLE 185
 DB 253 DADINGLRQVNDLTKMSDLEMYETLQELMA-----LKNKKEKESQLTGNS----- 303
 QY 186 VLVKVCCHKRRKDVSCPIROYPTGKQVPLIPDLNTPGPNFSLAVSNFEPSNSHMV 245
 DB 304 -----GDVNVETINAP--CKDLTKTLNDKRO---EYEQI-LAKNKKDIENQYET 346
 QY 246 KSYSLRLRYTPRGR-----REFNGMNGETNENIDVNEEL--PARRKRNEDEKTPVA 297
 DB 347 QITQIEHEVSSSGGVSSAKETVQLRHGVQLEIEIQLQSLSKRAALEKSLDTPKNRYCG 406
 QY 298 QMTVFDR---NRRQLLDGEYEVAMQEME-----ECPISKRATMETIDGKRLPPE 347
 DB 407 QLOMIQEDISNLEAQITDVREIECQNEYSLLSLIKMKLEKLEYTHNLLEGQ-EDFE 465
 QY 348 TFSQG 352
 DB 466 SSGAG 470
 RESULT 8
 CHD1_MOUSE STANDARD; PRT; 1711 AA.
 AC PA0201;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Chromodomain-helicase-DNA-binding protein 1 (CHD-1).
 GN CHD1 OR CHD-1.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-93211972; PubMed-8460153;
 RA Delmas V., Stokes D.G., Perry R.P.;
 RT "A mammalian DNA-binding protein that contains a chromodomain and an
 RT SNF2/SWI2-like helicase domain".
 RL Proc. Natl. Acad. Sci. U.S.A. 90:2414-2418(1993).
 CC -1- FUNCTION: SEQUENCE-SELECTIVE DNA-BINDING PROTEIN. COULD PLAY AN
 CC IMPORTANT ROLE IN GENE REGULATION.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: ABUNDANCE IS HIGHER IN CELLS REPRESENTING
 CC EARLY STAGES OF THE B LYMPHOID LINEAGE SUCH AS PRE-B AND B CELLS,
 CC THAN IN CELLS REPRESENTING MATURE PLASMACYTES OR OTHER CELL
 CC LINEAGES SUCH AS FIBROBLASTS.
 CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----


```

CC EMBL: L10410; AAB08486.1; -
DR HSSP; P23197; IAP0.
DR MGI: 88393; Chnl.
DR GO: 0005634; C:nucleus; NAS.
DR GO: 0003682; F:chromatin binding activity; NAS.
DR GO: 0006333; P:chromatin assembly/disassembly; NAS.
DR InterPro: IPR000953; Chromo.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR000330; SRF2_N.
DR Pfam: PF00385; Chromo; 2.
DR Pfam: PF00271; Helicase_C; 1.
DR Pfam: PF00176; SRF2_N; 1.
DR SMART; SM00298; CHROMO; 2.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR PROSITE; PS00598; CHROMO_1; 2.
DR PROSITE; PS00133; CHROMO_2; 2.
DR DNA-binding; ATP-binding; Helicase; Nuclear protein; Repeat.
KW DOMAIN 1 70 SER-RICH.
FT DOMAIN 116 136
FT DOMAIN 270 362 CHROMO 1.
FT DOMAIN 387 450 CHROMO 2.
FT NP_BIND 504 511 ATP (POTENTIAL).
FT SITE 612 615 DEAD BOX.
FT DOMAIN 1629 1645 3 x 5 AA REPEATS OF H-S-D-H-R.
FT REPEAT 1629 1633 1.
FT REPEAT 1635 1639 2.
FT REPEAT 1641 1645 3.
SQ SEQUENCE 1711 AA; 196409 MW; FE3F8F8D13E32E24 CRC64;

```

Query Match Best Local Similarity 4.0%; Score 153.5; DB 1; Length 1711; Matches 124; Conservative 80; Mismatches 207; Indels 205; Gaps 31;

```

QY 5 KKGGGGGGGSPSAGSGGSGGSAVAATAAGSGKSGCGGSGSYSSSSAAAAA 64
DB 11 RRGSGSSGSDGDCGASGSGSSGSSSDGSSQS -GSSDSDSGSDGSGSSSDPS 69
QY 65 GAALVPRKPKMHNQADHFLQAEKPTQIRFRTNLNLAIPILHRTLTYSIRNR 124
DB 70 RKNKQAKPKVADGAS-----FWKSSPILAVORSAMLRKPOAOQOPASSNSGEED 124
QY 125 TIKRRTFYVDMLSTVEKMGGEQESHSLAHQLFTFGFHNDKPS--PSENEONS 181
DB 125 SSSSEDS--DSSSSAKRKHNDKEDWQSG-----SGSPSOLGSSSESE 167
QY 182 VTLEVLIVVCHKKRKDVSCPIROVPTGKKOYPLIDLMOTKRGNFPSLAVSSNEFPSN 241
DB 168 -----ERDKSSCD-----GTESDEPRN 185
QY 242 SHWVKSYSLLFRTYRGR--EFNG-MINGENTENIDV-----NEELPRARRRN- 287
DB 186 K-----VSRKQRQNSKSKNGKILGQKRQIDSSSEDEDEDYDNDKSSRKQATV 236
QY 288 ----REDGE-KTFVQMTVFEDKRRRLQL-----DDEYEVAMQMEECPISSKRATV 334
DB 237 NVSYKDEDEMT-----DSDDLLEVCGEDVPEDEDEFE-TIRVMDCRGRGATG 287
QY 335 ERI-----LDGRRLPFEFTFSGGPIQLQTLKRTGTNDKSTAPAKPLATRNSESLHGE 388
DB 288 ATTITVAVADDDPNAGFERNKEPGDIOYLIRKKKSHHNT-----WETETLIKQO 339
QY 389 NRPGSVKTPTQTLAVKESLTLDQTRKEKTPN--ENRQRLRFYQGLYNNNTMQOTEAD 446
DB 340 NVRGMR-----LDNYKKADQETKRLKLNASEDEY--YN-----CQQLTD 380
QY 447 DLHCPWCLNCRILYSILKHLKICHSR-----FIENYVHPKGAIDVS 490
DB 381 DLH-----KQYQIVERI-IAHSNOKSAGLPLDYCKMGGLPSPSECSWEGALISKR 430
QY 491 INECYDGYAGNPQDIIHROPGRFASRNGPVKRTPI--THILVCRPRTKASMSSELESBD 548

```

```

DB 431 FQCID-EY-----FSRN-OSKTPFKCKVLAORPR-----FV----- 462
QY 549 GEVEOORTYSGGNRL 564
DB 463 -ALKOPSTIGCHEGL 477

RESULT 9
DSX_DROME STANDARD; PRT; 549 AA.
ID DSX_DROME
AC P23023; P23022; Q9VHY0;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Doublesex protein.
GN DSX OR CG11094.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM FEMALE AND MALE).
RC TISSUE=Larva, and pupae;
RX MEDLINE=89168451; PubMed=2493994;
RA Burtis R.C., Baker B.S.;
RT "Drosophila doublesex gene controls somatic sexual differentiation by
RT producing alternatively spliced mRNAs encoding related sex-specific
RT polypeptides."
RL Cell 56:997-1010(1989).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM MALE).
RC STRAIN=Barkeley.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer V.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burtis R.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson R., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fessler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Hock J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,
RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Moharry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Stryckas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).

```

RN [3]
 RP DNA-BINDING.
 RA MEDLINE-91330881; PubMed-1907913;
 RA Buttle K.C., Coschigano K.T., Baker B.S., Wensink P.C.;
 RA "The doublesex proteins of *Drosophila melanogaster* bind directly to a
 RA sex-specific yolk protein gene enhancer.";
 RL EMBL J. 10:257-258(1991).
 RN [4]
 RP DNA-BINDING DOMAIN, AND MUTAGENESIS.
 RA MEDLINE-93178426; PubMed-8440242;
 RA Erdman S.E., Buttle K.C.;
 RA "The *Drosophila* doublesex proteins share a novel zinc finger related
 RA DNA binding domain.";
 RL EMBL J. 11:527-533(1993).
 CC -1- FUNCTION: CONTROLS SOMATIC SEXUAL DIFFERENTIATION. BINDS DIRECTLY
 CC AND SPECIFICALLY TO THE ENHANCER FBE (FAT BODY ENHANCER) OF THE
 CC YOLK PROTEIN 1 AND 2 GENES (Y1 AND Y2). THIS ENHANCER IS
 CC SUFFICIENT TO DIRECT THE FEMALE-SPECIFIC TRANSCRIPTION
 CC CHARACTERISTIC OF THE Y1 GENES IN ADULT FAT BODIES.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named Isoforms-2:
 CC Name-Male:
 CC Isoform-23023-1; Sequence-Displayed;
 CC Name-Female:
 CC Isoform-23023-2; Sequence-VSP_001321, VSP_001322;
 CC -1- MISCELLANEOUS: EXPERIMENTALLY SHOWN TO BIND ZINC.
 CC -1- SIMILARITY: Contains 1 DM domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M25292; AAA17840.1; -;
 DR EMBL; M25293; AAA17841.1; -;
 DR EMBL; M25294; AAA17842.1; -;
 DR EMBL; AE003676; AAF54168.1; -;
 DR PIR; A32372; B32372.
 DR PIR; B32372; B32372.
 DR PDB; 1LPV; 02-OCF-02.
 DR TRANSFAC; T00955; -;
 DR TRANSFAC; T00956; -;
 DR FLYbase; FBgn0000504; dax.
 DR GO; GO:0003729; F:RNA binding activity; NAS.
 DR GO; GO:0003700; F:transcription factor activity; NAS.
 DR GO; GO:0007619; P:courtship behavior; NAS.
 DR GO; GO:0045487; P:female analla morphogenesis (sensu Holometabola); NAS.
 DR GO; GO:0007486; P:female genital morphogenesis (sensu Holometabola); NAS.
 DR GO; GO:0019101; P:female somatic sex determination; NAS.
 DR GO; GO:0007483; P:genital disc metamorphosis; NAS.
 DR GO; GO:0045436; P:female analla morphogenesis (sensu Holometabola); NAS.
 DR GO; GO:0007485; P:female genital morphogenesis (sensu Holometabola); NAS.
 DR GO; GO:0019102; P:female somatic sex determination; NAS.
 DR GO; GO:0007548; P:sex differentiation; NAS.
 DR InterPro; IPR001275; DM-DNA-binding.
 DR Pfam; PF00751; DM-domain; 1.
 DR SMART; SMO0301; DM; 1.
 DR PROSITE; PS40000; DM_DOMAIN_1; 1.
 DR PROSITE; PS50809; DM_DOMAIN_2; 1.
 KW Sexual differentiation; Alternative splicing; DNA-binding;
 KW Transcription regulation; Nuclear protein; Zinc; Metal-binding;
 KM 3D-structure.
 FT DNA_BIND 44 91 DM.
 FT DOMAIN 119 224 HIS-RICH.
 FT DOMAIN 267 296 SER/GLY-RICH.
 FT VARSPIC 398 427 ARVEINRTVAQIYYNYTPMALVNGAPMYL -> GQYVYNE
 FT YSRQHNINIVDGELNRTTRCG (In Isoform
 FT Female).
 FT /FTId-VSP_001321.

FT VARSPIC 428 549 Missing (In Isoform Female).
 FT FT
 FT MUTAGEN 47 47 /FTId-VSP_001322.
 FT MUTAGEN 50 50 C->A,H: ABOLISHES DNA-BINDING.
 FT MUTAGEN 59 59 H->Y: ABOLISHES DNA-BINDING.
 FT MUTAGEN 68 68 H->Y: ABOLISHES DNA-BINDING.
 FT MUTAGEN 70 70 C->D,Y: ABOLISHES DNA-BINDING.
 FT MUTAGEN 91 91 C->Y: ABOLISHES DNA-BINDING.
 FT MUTAGEN 91 91 R->Q: ABOLISHES DNA-BINDING.
 SO SEQUENCE 549 AA; 57409 MW; 3C1B92724E4CE083 CRC64;
 Query Match 3.9%; Score 153; DB 1; Length 549;
 Best Local Similarity 27.5%; Pred. No. 0.073;
 Matches 47; Conservative 27; Mismatches 47; Indels 50; Gaps 5;
 QY 11 GGGSGPSAGSGGGGGAATAATAGGKGGGSGGGSYSA----- 54
 DB 242 GGSVPATSSSGGAPSSNNAATSSNGSGGGGGGGGAGGGRSSGVITSA 301
 QY 55 -----SSSSAAAAGCAVLP-----VKKPKMEHYQADHEFLQAF 91
 DB 302 DHHMTVPTPAQSLGSCDSSSPSPSTSGAALIPISVVRKNGANVPLQGVFLDYCQ 361
 QY 92 KPTQYRFLRTNLIAPFLRTLYSHRSRTNIRKRTKVDMSKVE 142
 DB 362 KLEKFRY-----PHEL-MPLMTVILKDAADNIEASRIIE--ARVE 401
 RESULT 10
 K2CL_MOUSE
 ID K2CL_MOUSE STANDARD; PRT; 627 AA.
 AC P04104; 01-NOV-1986 (Rel. 03, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Keratin, type II cytoskeletal 1 (Cytokeratin 1) (67 kDa cytokeratin).
 GN KRT1 OR KRT2-1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RX [1]
 RP MEDLINE-85207740; PubMed-2581964;
 RA Steinert P.M., Parry D.A.D., Idler W.W., Johnson L.D., Steven A.C.,
 RA Rupp D.R.;
 RT "Amino acid sequences of mouse and human epidermal type II keratins
 RT of Mr 67,000 provide a systematic basis for the structural and
 RT functional diversity of the end domains of keratin intermediate
 RT filament subunits.";
 RL J. Biol. Chem. 260:7142-7149(1985).
 RN [2]
 REVISIONS, AND SEQUENCE FROM N.A.
 RA Rupp D.R.;
 RL Submitted (JAN-1999) to the EMBL/Genbank/DBD databases.
 CC -1- SUBUNIT: HETEROETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.
 CC -1- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
 CC MICROFILAMENT KERATIN, I (ACIDIC) AND II (NEUTRAL TO BASIC)
 CC (40-55 AND 56-70 KILODALTONS, RESPECTIVELY).
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M10937; AAD05191.1; -;
 DR SWISS-2DPAGE; P04104; MOUSE.
 DR MGD; MGI:96698; Krt2-1.
 DR InterPro; IPR001664; IF.
 DR InterPro; IPR003054; Keratin_II.

CC Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OK NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86302204; PubMed=3043197;
RA Gaber R.F., Styles C.A., Flink G.R.;
RT "PRK1 encodes a plasma membrane protein required for high-affinity
RT potassium transport in Saccharomyces cerevisiae.";
RL Mol. Cell. Biol. 8:2848-2859(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288c / FYJ679;
RA Czapluch C., Kordes E., Pujol A., Jauniaux J.-C.;
RT "Sequencing analysis of a 40.2 kb fragment of yeast chromosome X
RT reveals 19 open reading frames including URA2 (5' end), TRK1, PBS2,
RT PSPI0, GCPI4, REP1, PHO86, NCY3, ASF1, CCT1, GEF3, two tRNA genes,
RT three remnant delta elements and a Ty4 transposon.";
RL Yeast 12:1471-1474(1996).
CC -!- FUNCTION: THIS PROTEIN IS REQUIRED FOR HIGH-AFFINITY POTASSIUM
CC TRANSPORT.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M21328; AAA84728.1; -.
DR EMBL; Z49404; CAA89424.1; -.
DR PIR; S05849; PMBYH.
DR SGD; S0003665; TRK1.
DR GO; GO:0015079; F:potassium ion transporter activity; IDA.
DR GO; GO:0030007; P:potassium ion homeostasis; IDA.
DR InterPro; IPR003445; Cat_transp.
DR InterPro; IPR004773; ktransp_euk.
DR Pfam; PF02386; TrkH; 1.
DR TIGRFAMs; TIGR00934; Za38euk; 1.
KW Transports; Transmembrane; Potassium transport; Glycoprotein.
FT TRANSMEM 49 70 POTENTIAL.
FT TRANSMEM 78 98 POTENTIAL.
FT TRANSMEM 107 127 POTENTIAL.
FT TRANSMEM 778 800 POTENTIAL.
FT TRANSMEM 813 834 POTENTIAL.
FT TRANSMEM 838 858 POTENTIAL.
FT TRANSMEM 862 882 POTENTIAL.
FT TRANSMEM 898 918 POTENTIAL.
FT TRANSMEM 923 943 POTENTIAL.
FT TRANSMEM 971 991 POTENTIAL.
FT TRANSMEM 1078 1098 POTENTIAL.
FT TRANSMEM 1111 1131 POTENTIAL.
FT CARBOHYD 100 100 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 169 169 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 222 222 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 227 227 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 251 251 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 369 369 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 383 383 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 497 497 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 580 580 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 677 677 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 919 919 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 1030 1030 N-LINKED (GLCNAC . . .) (POTENTIAL).
FT CARBOHYD 1135 1135 N-LINKED (GLCNAC . . .) (POTENTIAL).
SEQUENCE 1335 AA; 141072 MW; BCSFEDA0BA09B2B CRC64;

	Query Match	3.8%;	Score 146.5;	DB 1;	Length 1235;
	Best Local Similarity	20.1%;	Pred. No. 0.47;		
	Matches	154;	Conservative	90;	Mismatches 227; Indels 295; Gaps 39;
QY	97	YRLTRNLAPLPIELHRTLYSMHRSNTFNIRKTFKVDMLSKV-----	-----	EKKM	145
Db		118	FKMRRTITLIERELTARTMTKRNRTGTOTSYPRKQAKDDDEQKLFSGEMVRNDEDSVH		207
QY	146	GEQSHSLISANLQLTLEFGF-----	-----	FHKNDKPSRPSNEQKSVYLEVLLVYCHKKR	196
Db		208	SDQSHSDLSRDSSNNNTNHNHNGSSGLDDFEVDEDEDDGGEQENK-----	-----	252
QY	197	KDVSCPIRQVPTGKROYPLIDNLQTKRGNPPSLAVSSNEE-EPSSNHMVYSILF---	---		252
Db		253	-----	-----	283
QY	253	-RYTRPGHRENGMHNGETNENIDVNEELPARKKRNRPDEKTYVAQMTVDKRRRLQL		311	
Db		284	HSKORPAP-----	-----	318
QY	312	DGEYVAMQMEECPI-----	-----	SKRATWETILDGKRLPPREFSQGPTL	355
Db		319	QGRHE-ATADEGCPPIVIGSPADGTRYKSNVKKLKAQG-----	-----	374
QY	356	QFTLRMTGETNDKSTAPIAKPLATRNSLSHOE-NKPGSVKPIQT-----	-----	IAYESL	406
Db		375	QNSV--SSEAN--STASVS-----	-----	423
QY	407	TTDLOTREKQTPPENRQKLIIFYOFLVNNTR--	---	QOTEARDDLHCWCTLNCKLVS-	462
Db		424	ETD-----	-----	468
QY	463	--LKLHLKCHSRLEFVYVHPKCARIDVSI-NBCYDGSYAGNODIHROPAPFSRNGP		519	
Db		469	KYLKKHF-----	-----	502
QY	520	VKRPPIHILVCRPKRTKAMSE-----	-----	FLSEDEGEVQORTYSSGHRL	564
Db		503	SDRPPII-----	-----	546
QY	565	YFSDTCLPLRPOEME-----	-----	VSEDEKDEPWLREKTIITOIEESDVNE	608
Db		547	--HSDSELKSHQOQOEKQHLQNLHMYKTSFSDNRSRAVPMERSRTI-DMAEKADLNE		603
QY	609	-----GKEVYMKLMLNH-----	-----	VAKHG-FLADONMHACMLFYENTQCK	646
Db		604	LARTPDFOKMYVQWKKAHRRKPNFRKRGNNKKIFEHGPIYVSDSRNPDI--	---	658
QY	647	IIRKNLCSNFMLHYSMDFMLISIMSDKAVYTKLREMQOKLEKGESAPANEETEEQN		706	
Db		659	-----NSLIHLAE-----	-----	695
QY	707	GT-----ANGFSEINSKEKA-----	-----	LETDSVSGVSKQSKOKL	739
Db		696	GSDHNGLNINPTYNDDDEGYGLHFDYDLDLDRHDSKSGKTYL		741
RESULT 13					
MSP1_PLAFW STANDARD: PRT: 1639 AA.					
AC	P04933:				
DT	13-AUG-1987 (Rel. 05, Created)				
DT	01-FEB-1996 (Rel. 33, Last sequence update)				
DT	15-SEP-2003 (Rel. 42, Last annotation update)				
DE	Mercozote surface protein 1 precursor (Mercozote surface antigens)				
GN	(P195).				
OS	MSP-1.				
OS	Plasmodium falciparum (isolate Wellcome).				
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
RN	NCBI_TaxID=5848;				
TX	SEQUENCE FROM N.A.				
TX	MEDLINE=86014355; PubMed=2995820;				

Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., Riveros-Moreno V.,
 RA Nicholas S.C., Hillman Y., Davey L.S., Tizard M.L.V., Schwarz R.T.,
 RA Freeman R.R.;
 RT "Primary structure of the precursor to the three major surface
 RT antigens of Plasmodium falciparum merozoites.";
 RL Nature 317:270-273(1985).
 RP REVISIONS.
 RA Holder A.A.;
 RL Submitted (Mar-1991) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 CC (Potential).
 CC -1- Ptm: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
 CC kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X02919; CA26676.1;
 DR PIR: A24594; A24594.
 DR PIR: S05603; S05603.
 DR PDB: 1CEJ; 28-MAY-99.
 DR InterPro: IPR006209; EGF_Like.
 DR Pfam: PF00008; EGF_1.
 KM Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 KM Transmembrane; GPI-anchor; 3D-structure.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1639 MEROZOITE SURFACE PROTEIN 1.
 FT CARBOHYD 116 116 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 268 268 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 764 764 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 768 768 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 783 783 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 844 844 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 920 920 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 964 964 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1058 1058 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1165 1165 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1174 1174 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1445 1445 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1526 1526 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 1639 AA; 187618 MW; 2C255B6616C87F6E CRC64;
 Query Match 3.8%; Score 146.5; DB 1; Length 1639;
 Best Local Similarity 20.6%; Pred. No. 0.67;
 Matches 168; Conservative 123; Mismatches 291; Indels 235; Gaps 46;
 QY 10 GGGGSGPSAGSGG-GFGSAVAAMATAGGKSGGSGGSGGSSSSSSAAAAAGAAV 68
 DB 68 GSKGSVASGSGSVASGSGSVASGSGSVASGSGSNSRRTSPDSNDSPDASYS- 126
 QY 69 LPPVKPKMEVADHFLQAFEPKPTQIYRFLRRLAPILFLHRLTYMSHRNSNTNK 128
 DB 127 -----DLKHYRNYLTITKELKYP-QLDDL-TINML-----TLG-----DNTH 162
 QY 129 RTFFVADMLSKVEKKKGESHSLSAHLQLTFTEGFHNRKPSNSENDONSVTLEVL 188
 DB 163 GEFYLD-----GEEINELLYKLNFEYD-----LIRAKL 192
 QY 169 VVYCHKKRDVSCPIRQVPTGKKQVPLIDLNOTKGNPPLASVSNFEPSPSHAVKSY 248
 DB 193 NDVC-----ANDYC-----QIPF-----NKRINELD-----VLK-- 218
 QY 249 SLLEFVTRP-----GRREFNGMNGETENIDVNEELPARRKRNREDGEKTFVAQMT 300
 DB 219 KLVFGYRKLNDIKNDVGNKMDYIKNNKTIENI--NELI-----EESKTIQKNKN 268

QY 301 VEDKRRRLQLDGEYEVAN--QEMEC-----DISKRATWETILDG---KRL 343
 DB 269 ATKREERKKLYQAOYDLSTYNNKOLEBANHLISYLERIDPLTKNENIKFELDKINEKNP 328
 QY 344 PPEPFPSCPTQFLQFLMNGEINDSTAIAPLATRNS-----LHQENKKG 392
 DB 329 PPANS-GNTPNLLDKNNKIEEHEKEIKETATIF-NIDSFTDPLELEYLREKN-- 384
 QY 393 SVKPTQIAVKESLTDTLQTRKKDTPNPNRKLRIFYQFLVN--NNTROQTEARDLHC 450
 DB 385 -----NIDISAKVEI-----KESTPNENPN--GYTYLSINDINNALNELNSFDLLN 431
 QY 451 PW-----CTINCRFLY--SLKHLKLSHRETFN--YVHPGARI-DVS----- 490
 DB 432 PDYTKEPKSNITTDNERKKFLINEIKERIKIEKKIESDKSIEDRSKSLNDITREKYL 491
 QY 491 INECYGSAGNPO--DIHQGFAPSRNGPKRPITITILVCRPKRTASSEF---LE 545
 DB 492 LNEIYDSKFNNDLTNFERKMGKRS---YKVKLTG-----HNTFASYENSKHNE 541
 QY 546 SEDGEVQOQRTYSSG-----NRLYFSDTCLPLRQEMEVSDP-EKDEPMLEKTIQI 600
 DB 542 KLTALKIYEDYSLRNIVVEKELKYKNLSKIE-NEITLYENIKKDEQLPEKTYR- 599
 QY 601 EEFSDVNEGEKEVEMKLMN-HVAKHGFIADONNHNACMLEVYGGRIIKKLCRMPLH 659
 DB 600 -----DENKPKDEKLEVSIVKQVQVLLMNKID-----ELKKTQLLKNVELKHNH 648
 QY 660 LVSMH-----DENISIMSDKATYKLRMOOKIE-----KESSAPANEET 702
 DB 649 VPSYVQENKQKEPYLLI---VLKKEIDKLVMPYVESLINEKNKIKTEGSDNSEPST 705
 QY 703 EEOGTFANGFSEINSREK---ALETDSVGSVSKOSK 736
 DB 706 E---GEITGOATTKRPGQAGSALGSDSVQAQADEQO 739
 RESULT 14
 GOG4_HUMAN
 ID GOG4_HUMAN STANDARD; PRT; 2230 AA.
 AC Q13439; Q13270; Q13654; Q14436;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Golgi autoantigen, golgin subfamily A 4 (Trans-golgi p230) (256 kDa
 DE golgin) (Golgin-245) (72.1 protein).
 GN GOLGA4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RX MEDLINE=96215236; PubMed=8626529;
 RA Erlich R., Gleeson P.A., Campbell P., Dietzsch E., Toh B.-H.;
 RT "Molecular characterization of trans-golgi p230: a human peripheral
 RT membrane protein encoded by a gene on chromosome 6p12-22 contains
 RT extensive coiled-coil alpha-helical domains and a granin motif.";
 RL J. Biol. Chem. 271:8328-8337(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RL Seelig H.P.;
 RN Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
 RP [3]
 RP SEQUENCE OF 131-2230 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=96125112; PubMed=8537393;
 RA Fritlier M.J., Lung C.-C., Hamel J.C., Griffith K.J., Chan E.K.L.;
 RT "Molecular characterization of golgin-245, a novel Golgi complex
 RT protein containing a granin signature.";
 RL J. Biol. Chem. 270:31262-31268(1995).
 RN [4]
 RP SEQUENCE OF 524-672 FROM N.A.

Qy	18	SVLLEVLVYVCHKKRRKDVSCPIROVPGKKVPLPIPLNQTOKGNPISLAVSSNEEPS	240
Db	181	SVLLEVLVYVCHKKRRKDVSCPIROVPGKKVPLPIPLNQTOKGNPISLAVSSNEEPS	240
Qy	241	NSHMYVSYSLLFEVTPPGRRFNGMINGETNENIVNEELFARRKRRNDEKTFVQMT	3000
Db	241	NSHMYVSYSLLFEVTPPGRRFNGMINGETNENIVNEELFARRKRRNDEKTFVQMT	3000
Qy	301	VDFKRRRLDLDBGEYVAVQMEBCPISKRRATWETIUDGRLLPPEFTFSOGPTLOFTLR	3600
Db	301	VDFKRRRLDLDBGEYVAVQMEBCPISKRRATWETIUDGRLLPPEFTFSOGPTLOFTLR	3600
Qy	361	WQETNDKSNAPLAKELATRNSESLHOENKESVAPQOTIVKSSLTDTDOTREKXTPN	4200
Db	361	WQETNDKSNAPLAKELATRNSESLHOENKESVAPQOTIVKSSLTDTDOTREKXTPN	4200
Qy	421	ENROKRLIFYOLFYNNTROOTEARDDLHCPCSTLNCRLKLSLKHKLCHSRIFMYVY	4800
Db	421	ENROKRLIFYOLFYNNTROOTEARDDLHCPCSTLNCRLKLSLKHKLCHSRIFMYVY	4800
Qy	481	HPGARGAIDVSYNECYDGSYAGNPODIHHQPGFASRNGPVKRTPTIHLVCRPKRTASM	5400
Db	481	HPGARGAIDVSYNECYDGSYAGNPODIHHQPGFASRNGPVKRTPTIHLVCRPKRTASM	5400
Qy	541	SEPLESEDDVEGOQRYSGGHNRLKLFHSDTCLPLRPMEMVSDSEKDEPELAKRTTQI	6000
Db	541	SEPLESEDDVEGOQRYSGGHNRLKLFHSDTCLPLRPMEMVSDSEKDEPELAKRTTQI	6000
Qy	601	EEFSDVNEGEKEVMKLMNLHVAKHGFIADNQNHACMLFEVNYOSOKIIKKMLCENFMJHL	6600
Db	601	EEFSDVNEGEKEVMKLMNLHVAKHGFIADNQNHACMLFEVNYOSOKIIKKMLCENFMJHL	6600
Qy	661	VSHHDFNLISIMSIDKAVTKLEMOOKLEKESASPANEITEBQNTANGFSEINSKEK	7200
Db	661	VSHHDFNLISIMSIDKAVTKLEMOOKLEKESASPANEITEBQNTANGFSEINSKEK	7200
Qy	721	ALETDSVSGVSKOSKROKL 739	
Db	721	ALETDSVSGVSKOSKROKL 739	

RESULT 2
US-09-874-162A-8
Sequence B Application US/09874162A
; Patent No. US20020155452A1

APPLICANT: Koonitz, Jason
APPLICANT: Sklar, Jeffrey
TITLE OF INVENTION: FUSION OF JAZF1 AND JAZF1 GENES IN
TITLE OF INVENTION: ENDOMETRIAL STROMAL TUMORS
FILE REFERENCE: 05311-024001
CURRENT APPLICATION NUMBER: US/09/874,162A
CURRENT FILING DATE: 2001-06-04
PRIOR APPLICATION NUMBER: US 60/209,093
PRIOR FILING DATE: 2000-06-02
NUMBER OF SEQ. ID NOS: 23
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8
LENGTH: 776
TYPE: PRF
ORGANISM: Homo sapiens
US-09-874-162A-8

Query Match	87.9%	Score 3413	DB 10	Length 776
Best Local Similarly	99.8%	Pred. No. 6.9e266		
Match 647; Conservative	1	Mismatches 0	Indels 0	Gaps 0

QY 92 KPTQIRPLRTNLAPILFHTLTLYMSHRNRTIKRKTQVDMLSKYEKMGDESH 151
:|||||
Db 129 EPTQIRPLRTNLAPILFHTLTLYMSHRNRTIKRKTQVDMLSKYEKMGDESH 188
:|||||
QY 152 SLSAHQLFTGTFHNDKPSNSEQNSVTLFVLLKVCCHKRRDYSCITRVPYGGK 211
:|||||

Dd	189	SLSAHLQJLTFTGFFPHKHNKDRSPNSENEQNSVTLLEVLVAVCHHKRRDVOSPIRQVPTGK	248
Qy	212	QVPLIPDLNQTKPAGNPSLAVSNEFEPSNSHMVKSYSLLFRVTRGRRENFNGINETN	271
Dd	249	QVPLIPDLNQTGPNFNPPLAVSNEFEPSNSHMVKSYSLLFRVTRGRRENFNGINETN	308
Qy	272	ENIDVNEELPARKKNNREDGEKTFVAQOMVFPQKNRRLQJLDGXYEVAQOMECEPTISKRR	331
Dd	309	ENIDVNEELPARKKNNREDGEKTFVAQOMVFPQKNRRLQJLDGXYEVAQOMECEPTISKRR	368
Qy	332	ATWETIILDGKRLLPPEETFSQGPJTLOFTLRWGTGETNDSAPLAKPLATRNSESLHOENKP	391
Dd	369	ATWETIILDGKRLLPPEETFSQGPJTLOFTLRWGTGETNDSAPLAKPLATRNSESLHOENKP	428
Qy	392	GSVRFQTIQAVKESLTTDQJTRKEKDTPNENOKRLFIYQFLYNNNTROOTEARDDLHCP	451
Dd	429	GSVRFQTIQAVKESLTTDQJTRKEKDTPNENOKRLFIYQFLYNNNTROOTEARDDLHCP	488
Qy	452	WCTLWCRKLYSLKLHLKLCHSRPIFYVYVHPRGARIDVINSICYDOSYAGNPODIHROP	511
Dd	489	WCTLWCRKLYSLKLHLKLCHSRPIFYVYVHPRGARIDVINSICYDOSYAGNPODIHROP	548
Qy	512	FAFSNGPVRKTPITHILVCRPKRTKASSEFLSESDGEVEOORTYSOGHNRLYFHSPTC	571
Dd	549	FAFSNGPVRKTPITHILVCRPKRTKASSEFLSESDGEVEOORTYSOGHNRLYFHSPTC	608
Qy	572	LPLRPOEMEVDSEDEKDEPMLREKTTQJTEEFSDVNEGEKEVWKLNNLHYMKHGFADNQ	631
Dd	609	LPLRPOEMEVDSEDEKDEPMLREKTTQJTEEFSDVNEGEKEVWKLNNLHYMKHGFADNQ	668
Qy	632	MNHACMLFVNTGOKIILKKNLCRNFMHLVSMHDFNLISIMSDIKAVTKLREMOQKLEKG	691
Dd	669	MNHACMLFVNTGOKIILKKNLCRNFMHLVSMHDFNLISIMSDIKAVTKLREMOQKLEKG	728
Qy	692	ESASANEITEEONGTANGFSEINSKEALETDVSQVSKOSKOKRL	739
Dd	729	ESASANEITEEONGTANGFSEINSKEALETDVSQVSKOSKOKRL	776

RESULT 3
US-09-764-864-931
; Sequence 931, Application US/09764864

```

: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: P1223
: CURRENT APPLICATION NUMBER: US/09/764,864
: PRIOR FILING DATE: 2001-01-17
: Prior application data removed - consult PALM or file wrapper
: NUMBER OF SEQ ID NOS: 1792
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 931
: LENGTH: 388
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-764-864-931

```

Query Match	52.68	Score 2043	DB 10	Length 388
Best Local Similarity	99.58	Pred. No. 4.5e-156		
Matches 386, Conservative	0	Mismatches 2	Indels 0	Gaps 0

Ky 52GFLQDLLENISGLNDKSLAFIAAPFLAIIINSESLQELNFGSVAFIQLIAVRESLTIDQ 411
 Db 1 GPTLOFTLMTGETNDKSTAPIAPPLATRNSESLHQENKPGSVKPTQIIAVKESLTIDQ 60

Accession	Protein	Length
57		47
58		47
59		47
60		47
61	TRREKDPNNENRQKLFIFQFLYNNNTROTAREARDLDCPMTCTLCRLYSILNHLKICH	120
D6		47
Q7	472 SRIFVNVVHPKARIDVINSNECYOSVAGNQNODIHRPGFAFSNGVKKPTPTTHLYC	533
D8	121 SRIFVNVVHPKARIDVINSNECYOSVAGNQNODIHRPGFAFSNGVKKPTPTTHLYC	180

NAME/KEY: SITE
LOCATION: (4)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (136)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (166)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1355

Query Match 23.7%; Score 919; DB 10; Length 175;
Best Local Similarity 94.3%; Pred. No. 4.1e-66;
Matches 165; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 429 FYQFLYNNNTROQTEARDDLCPCWTLNCRKLYSLKHLKLSRFIFNYHYHKGARID 488
DB 1 FYQFLYNNNTROQTEARDDLCPCWTLNCRKLYSLKHLKLSRFIFNYHYHKGARID 60
QY 489 VSINECYDGSYAGNPQDIHROPGFASRNGPVKRTPTIHLVCSRPKRTKASMSFLESSED 548
DB 61 VSINECYDGSYAGNPQDIHROPGFASRNGPVKRTPTIHLVCSRPKRTKASMSFLESSED 120
QY 549 GEVQDQRTSSGCHRLYFHSPTCLPRQEMEVSEDEKDEPWLKXKPLKRLKSF 603
DB 121 GEVQDQRTSSGCHRLYFHSPTCLPRQEMEVSEDEKDEPWLKXKPLKRLKSF 175

RESULT 7

US-10-231-778-2
Sequence 2, Application US/10231778
Publication No. US20030126647A1
GENERAL INFORMATION:
APPLICANT: Bildeau, Pierre
APPLICANT: Chaudhury, Abdul M.
APPLICANT: Dennis, Elizabeth S.
APPLICANT: Koltunow, Anna M.G.
APPLICANT: Luo, Ming
APPLICANT: Peacock, William J.
TITLE OF INVENTION: Method for inducing seed development by down-regulating
FILE REFERENCE: 72-98A
CURRENT APPLICATION NUMBER: US/10/231,778
CURRENT FILING DATE: 2002-11-08
PRIOR APPLICATION NUMBER: 09/398,237
PRIOR FILING DATE: 1999-09-20
PRIOR APPLICATION NUMBER: 60/101,184
PRIOR FILING DATE: 1998-09-21
PRIOR APPLICATION NUMBER: AU PP6061
PRIOR FILING DATE: 1998-09-22
PRIOR APPLICATION NUMBER: AU PP6062
PRIOR FILING DATE: 1998-09-22
PRIOR APPLICATION NUMBER: AU PP6063
PRIOR FILING DATE: 1998-09-22
PRIOR APPLICATION NUMBER: AU PP6064
PRIOR FILING DATE: 1999-07-01
PRIOR APPLICATION NUMBER: AU PP6065
NUMBER OF SEQ ID NOS: 239
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 2
LENGTH: 813
TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-10-231-778-2

Query Match 5.6%; Score 219.5; DB 15; Length 813;
Best Local Similarity 18.5%; Pred. No. 7.6e-09;
Matches 150; Conservative 115; Mismatches 306; Indels 239; Gaps 29;

QY 77 EHVQDHLFLQAFKPKQIYRFLRLNLIAPIFLHRTLYM--SHRNSRNIRKR-TFK 133
DB 57 DDVVDENIT--KTKKPAVYKTKLETRSKNNPYFLRLSLKTYIQAKKKKSSGGKIRFN 114

QY 134 VDDMLSKVEKMKGEQSHS-----LSAHL----- 157
DB 115 YRDVSNMKTAKAEVYENFSCPECLIPCGHGGQLHLKSSHDAKFEFYRAEKDHGPEVD 174
QY 158 -----QLETFGFFHKNDKPSNEBONSVTLEVLVYK----- 191
DB 175 VSVKSDIKRQVGLKDDVGNQPLSPFCSSKNRQORRDSNNVKKLNTVLMELDLDLP 234
QY 192 -----CHKRRKDVSCPIR-----QVDTGKQVPLIPDLNQTGRNF--P 228
DB 235 RGENDSTHVDDNVSPRAHSEKISDLITTTQLIAASSEKRVHND--GNVSSP 291
QY 229 SLAVSSNE-----FEPSNSHMYKSYSLP-----RYTPRRRFRNGMINEYN 271
DB 292 PRAHSSAKENSTHVNDODDVSPRAHSSLEKNESTHVNDONISPPKAH-----SSKN 346
QY 272 ENIDVNEE--LPARRNRNEDGKTYVAQTVYDK-----NRLQLLDGEYVAMQ 320
DB 347 ESTHMDNEDVSFPRTSSKETSDILTTTPAIVEPEPKVRGCRKOLYAKRY----- 401
QY 321 EMECPISKRTATYETIILDGKRLPPF-----ETFSQGP-----TLOFTLMTG 363
DB 402 -----KANETPALAESSEPKVYLHVNDENNSPPRAHSLKASDILTTTPALAES 453
QY 364 E-----TNDKSTADIAPLATRNSESLHQ--ENKPGSVKPTQTAVERESLTTDQTRKK 416
DB 454 EPKVPYHNDENVSTPRAHSSKKNKSTRKNVNDVSPPKTRSSKTSJDLITTTQPTAES 513
QY 417 DTP-----NENRKLRFYQFLYNNNTROQTEARDDLCPCWTLNCRKLYSLKHLKLC 470
DB 514 SEPKVRHVNDVNSTPRAHSSKKNKSTRKND--DNIPSPKTRSSKTSJDLITTTQPTAES 568
QY 471 HSRFTFNVYHPRKARIDVSINECYDGSYAGNP-----QDIHROPGFA----- 513
DB 569 -----PALAESEPKVPHVNDKVSSTPRAHSSKKNKSTRKND--DNIPSPKTRSSKTSJDLITTTQPTAES 617
QY 514 -----FSRNGPVKRTPTIHLVCRPKTKASMSFLESSEGEVEQDQRTSSGCHRLYF 567
DB 618 KTSJDLITTTQPTAES-----SEPKVTRSRKRELAERCEAKRLERLK--GRQFYH 668
QY 568 SDTCLPLRQPM--EVSEDEKDEPWLREKTIITIEERSDVNEBEKVMKMLNLYK 625
DB 669 SOTQPMPTFEQVWGNSENETDYALDISRLRLVGSKEKREMYLMNLFVKOR 728
QY 626 FIADNMHACMLEVE-----NYGOKIITKKNLCNFMHLVSMHDFILSIS 673
DB 729 VIADGHPWACEEFAKLKHEKMSSTFDMMWRFRFKIMNNGLICAKTFKCTTILSN 788
QY 674 IDKAVTKLREMOQKLEKESASPANEEITE 703
DB 789 SDEA-----GQFTSGSAAANNNOQSME 810

RESULT 8

US-10-097-340-35
Sequence 35, Application US/10097340
Publication No. US20030087250A1
GENERAL INFORMATION:
APPLICANT: John MORAHAN
APPLICANT: Manjula GANNAVARAPU
APPLICANT: Sebastian HOERSCH
APPLICANT: Shubhangi KAMATKAR
APPLICANT: Steve G. KOVATS
APPLICANT: Rachel E. MEYERS
APPLICANT: Michael MORRISSEY
APPLICANT: Peter OLANDT
APPLICANT: Ami SEN
APPLICANT: Peter VEIBY
APPLICANT: Gordon B. MILLS
APPLICANT: Robert C. BAST, Jr.
APPLICANT: Karen LU
APPLICANT: Rosemarie SCHMANT


```

0Y      8 GGGGGGGSPSSGGGGGGGSAVAALAAASGKSGC---GSGCGGGSVSASSSSSSAAAAA 64
      84 GGGGGGAGAGVAGVAGGGGSGA-----SSGGAGGLPPSSAGRPSPSPS----- 132
0Y      65 GAAVLPVKKPMEHQADHELELAQEFKPTQIYRELRLRNLAPIFELHRTLYMSHRNS 124
      133 --VSEKKELELERLQKEE-----ERKKRLDY-----VFWRCLAVPPNAPQ 175
0Y      125 TNIRKRTFFKVDMLSKYEK-----MKGEQESHLSLHLLQ---TTFGFPHKDKDSPN-- 174
      176 TDMARROOKISKQOLOTQVFKDFEAFNLETQIMDEAFMNAVOSYEVELFKSDRAVMQ 235
0Y      175 -----SENQNSVTLEVLVYCHKKRKRVSCPRIQVPPGKQVPLIPDLNQKPKGNFSL 230
      236 SGGCSANDSRETF-----KKHJ-----EKRVSLPELDIGSKETVLS 273
0Y      231 -----AVSNEFER-----SNHMYKSYLEFRVTRP---GREFNGMI--NGET 270
      274 WMAKFDATYREEDPRQQAARMTASAABELLSKQALYEMPONIGIKFFHQLLYMQ 333
0Y      271 NENID-----VNEELPAR--RKRNRDEKTFYAQMTVFDKNRRLDLDGEYVAMOEEMEE 324
      334 LDNDEQAQAQIRRELDGRLOMADQIARERKRPKFFKSKEMENYIELKSSVMILANLES 393
0Y      325 CRISK-----KRATWETLLOKRLRPREFESQPTLOTLTWT---GENDKSTAP- 372
      394 MPVSIGGEFKLOKLKRSHNASTIDGSESENOUSKSDVSLTSELVIMEVQGLSLAN 453
0Y      373 --IAKPLATRSESLH---OENKP-----GSVKPTQTI-AKESL---TTDLQTRKED 417
      454 RIVYCTMEVEGGEKQLODAESKRTWTQOGDFTTHALPAKVKYLFTFESTIVALLDEKE 513
0Y      418 -----TPENROKLRIFYOFLYNNNTROOTTEARDLHCPCWCTLNCRLKLYSLKHKL 469
      514 LGRVLIHPTPNSPKO-----SEWHKATVSK---NCPDOLKIKILAVMRKPOHN 559
0Y      470 CHSRIF---NYVYPRKCARIDVSIINECYDGAACNPQDINHQPAPFARNPVPARTPT 526
      560 KHSGLYALGNKVMKMKKREFVLQV-----SOYFA----- 592

```

QY 233 SSNEEPSNSHMKSYSLFF

	Query Match	3.9%	Score 151;	DB 15;	Length 669;
	Best Local Similarity	19.0%;	Pred No. 0.0019;		
	Matches 145;	Conservative 94;	Mismatches 228;	Indels 298;	Gaps 35
OY	5 KHGGGCGG-----GSGPSAGSGGGGFGGSAVAATASGKSGGGSCGGGSSYSASSSS 58				
Dd	2 EOGGGGGGNEVEEASPISSRPANNLLELMFSAADGDGAGGGGGGGGSSASSSSGN 61				
OY	59 SAAAAAGAAYLPVKPKFKHEHVADHFLQAEFKPQIQRILTNRNLAPICLH--RTLT 116				
Dd	62 RWPREETLLALNIR-----SDMDSTPRDALTILKPRLWEHVSRKLL 100				
OY	117 YMSHSNSRNTLRKRTEFYDDMLSKYEKKMGDESHSLAHULTFTGFPHK----NDKPS 172				
Dd	101 ELGYTRSSCKCKKEKENVOKYYTKRKRETRGR--HDGRKY-----KFSSOLEALNTTP- 151				
OY	173 PENSEONSVTLEVLIVVCCHRRKRDVSCPRIOVPTGKKOVPILIPLDLNQTGKNPSPLAY 232				
Dd	152 PSS-----SLDVNPISLVAN-----PILMPSSSSSPFVV---FSQPOP----- 185				
OY	233 SSNEFEPSNHVAKSYSLLFRTRP-----GRRENQINGENTNEN 273				

```
Db 186 -QOTOTPPQTHNVS-----FTPTTPTPLPLPSMGPFTGVTFFSSHSSTASGMSDDDD 239
Qy 274 IDVNEELPA-----RRKRRNREDEKTFVAMTVFD-----KNR 306
Db 240 MDVDAQANIAGSSSRKRRGNRGSGGK-----MMELFEGLVNQVMQKAAQMSFLEALEKR 295
Qy 307 RLQLLDGEYVAMQEM-----EECPISKRA-----TWETIIDGKRLLP 345
Db 296 EOERLDREEMAKRQEMARLAREREHVMSQERAAASRDAAIISLQKITGHTI-----OLPP 351
Qy 346 FETPSQPTLOFTLWMTGETNDKSTAPIAK-PLATRNS-----384
Db 352 SLSSQPPPTQ-----PPPAVTKRYAEPPLSTASQSQOPIMAIPQOILPPPSH 403
Qy 385 --LHQENKPGSVKPTQTIAYKE-----SLTTDLQTRKEKDP-- 419
Db 404 PHAQPOKQOQOQPOQEMVMSSEQSLPSSSRMPKAEIILINLSGMEPRYQDNVPGKL 463
Qy 420 --NENRQKRLIFQFLNNNTROOTEARDDLHCPCWCLNCRKLYSLKH-----LKL 469
Db 464 LMEIISTSMK--RMGYNRAKRCKE-----KWNENIN--KYYKKVESNKKRPODAKT 511
Qy 470 CHSRFIFNYVYHPKARIDVSIKCYDGYAGN-----PODIHQPGFAFSRNGPVKRT 524
Db 512 C-----PFHRLDLIRKXKVLGSGGSGSTSLPD-----OKOSP 546
Qy 525 ITHILVCPKRTKASMESEFLESEDEGEVEQOQRTYSSGHNRLYFHSDTCLPLRQEMEYDSE 584
Db 547 VT--AKKP-----PQGLVNVQOTHSASTE-----EEPIEESPQ 580
Qy 585 DEKPEWLREKTTQ-----TEFSDVNGEKE 612
Db 581 GTEKPEDLVNKLQOQOQLOQOESMIGYEKIEESHNNMEE 625

RESULT 10
US-10-278-536-12
: Sequence 12, Application US/10278536
: Publication No. US2003013186A1
: GENERAL INFORMATION:
: APPLICANT: Samaha, Raymond
: APPLICANT: Heard, Jacqueline
: APPLICANT: Jiang, Cai-Zhong
: APPLICANT: Pineda, Omaira
: APPLICANT: Reuber, Lynne
: APPLICANT: Riechmann, Jose-Luis
: APPLICANT: Yu, Guo-Liang
: APPLICANT: Keddie, James
: APPLICANT: Ratcliffe, Oliver
: APPLICANT: Pilgrim, Marsha
: APPLICANT: Adam, Luc
: APPLICANT: Brown, Pierre
: TITLE OF INVENTION: STRESS-INDUCED POLYNUCLEOTIDES
: FILE REFERENCE: MBI-011
: CURRENT APPLICATION NUMBER: US/10/278,536
: CURRENT FILING DATE: 2002-10-22
: PRIOR APPLICATION NUMBER: 60/125,814
: NUMBER OF SEQ ID NOS: 238
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 12
: LENGTH: 669
: TYPE: PRT
: ORGANISM: Arabidopsis thaliana
: FEATURE:
: OTHER INFORMATION: 6634
US-10-278-536-12

Query Match 3.9%; Score 151; DB 16; Length 669;
Best Local Similarity 19.0%; Pred. No. 0.0019;
Matches 145; Conservative 94; Mismatches 228; Indels 298; Gaps 35;
Qy 5 KHGGGGG-----GSGPSAGSGGGGFGSAAVAATAAGSGKSGGGGCGGGSISASSSS 58
```

```
Db 2 EOGGGGGGNEVEEASPISSRPANNLEELMRFSAAADGGLGGGGGSGSSSSGN 61
Qy 59 SAAAAAGAAVLPVKKPMENHVQADHLEFLQAFKRPQIYFLTRNLIAIFLH--RTL 116
Db 62 RMPREETLALLR-----SDMSTFDATKALPLMEHVSRL 100
Qy 117 YMSHRNRTNIRKRTKFEVDDMLSKVERMKGEOSHSIAHLQTLFTGFHK-----NDKPS 172
Db 101 ELGYKRSKCKCKEFENVQVYKRTKTRGCR--HOKAY-----KFSQLALNTTP- 151
Qy 173 PNSENEQNSTLEVLYKVCCHKRKDYSCIRQVPGKQKQVPLIPDLNQTGPNFSLAV 232
Db 152 PSS-----SLDTPLSVAN-----DILMPSSSSSPFV--FSQOP- 185
Qy 233 SSNEFSPNSHMVKSYSLLPRTVTP-----GREFPMINGENEN 273
Db 186 -QOTOTPPQTHNVS-----FTPTTPTPLPLPSMGPFTGVTFFSSHSSTASGMSDDDD 229
Qy 274 IDVNEELPA-----RRKRRNREDEKTFVAMTVFD-----KNR 306
Db 240 MDVDAQANIAGSSSRKRRGNRGSGGK-----MMELFEGLVNQVMQKAAQMSFLEALEKR 295
Qy 307 RLQLLDGEYVAMQEM-----EECPISKRA-----TWETIIDGKRLLP 345
Db 296 EOERLDREEMAKRQEMARLAREREHVMSQERAAASRDAAIISLQKITGHTI-----OLPP 351
Qy 346 FETPSQPTLOFTLWMTGETNDKSTAPIAK-PLATRNS-----384
Db 352 SLSSQPPPTQ-----PPPAVTKRYAEPPLSTASQSQOPIMAIPQOILPPPSH 403
Qy 385 --LHQENKPGSVKPTQTIAYKE-----SLTTDLQTRKEKDP-- 419
Db 404 PHAQPOKQOQOQPOQEMVMSSEQSLPSSSRMPKAEIILINLSGMEPRYQDNVPGKL 463
Qy 420 --NENRQKRLIFQFLNNNTROOTEARDDLHCPCWCLNCRKLYSLKH-----LKL 469
Db 464 LMEIISTSMK--RMGYNRAKRCKE-----KWNENIN--KYYKKVESNKKRPODAKT 511
Qy 470 CHSRFIFNYVYHPKARIDVSIKCYDGYAGN-----PODIHQPGFAFSRNGPVKRT 524
Db 512 C-----PFHRLDLIRKXKVLGSGGSGSTSLPD-----OKOSP 546
Qy 525 ITHILVCPKRTKASMESEFLESEDEGEVEQOQRTYSSGHNRLYFHSDTCLPLRQEMEYDSE 584
Db 547 VT--AKKP-----PQGLVNVQOTHSASTE-----EEPIEESPQ 580
Qy 585 DEKPEWLREKTTQ-----TEFSDVNGEKE 612
Db 581 GTEKPEDLVNKLQOQOQLOQOESMIGYEKIEESHNNMEE 625

RESULT 11
US-10-087-464-10
: Sequence 10, Application US/10087464
: Publication No. US20030039436A1
: GENERAL INFORMATION:
: APPLICANT: Chishtli, Athar
: APPLICANT: Oh, Steven
: APPLICANT: Liu, David
: APPLICANT: Goel, Vikas
: APPLICANT: Li, Xuerong
: TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses There
: FILE REFERENCE: S1237/7019
: CURRENT APPLICATION NUMBER: US/10/087,464
: CURRENT FILING DATE: 2002-03-01
: PRIOR APPLICATION NUMBER: US 06/272,930
: NUMBER OF SEQ ID NOS: 59
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 10
: LENGTH: 1639
: TYPE: PRT
```

ORGANISM: Plasmodium falciparum
US-10-087-464-10

Query Match 3.8%; Score 146.5; DB 15; Length 1639;
Best Local Similarity 20.6%; Pred. No. 0.016;
Matches 168; Conservative 123; Mismatches 291; Indels 235; Gaps 46;

```

OY 10 GGGGSPAGSGGCG-GEFGSAVAATAATAGSGGSGCGGSGYSASSSSAAAGAAV 68
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 68 GSKGSVAGSGSGSVAGSGSVAGSGSVAGSGSVAGSGSGNSHRTNPSDSSDAKSYA- 126
OY 69 LPVKPKMEHQADHEFLQAEPKPTQIYRFLRTNLNLAIPILHRTLTMTMSRNSNTNK 128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 127 -----DLKRVNRLTLTKELKYP-QLFDL--TNHML-----TLC-----DNH 162
OY 129 RKTFFVDMLSKEVKKEGEOESHSLAHQLFTFGFFHNDKRPSPNEQNSVTEVL 188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 163 GKYLID-----GVEINELLYKLNIFYD-----LIRAKL 192
OY 189 KVCHKKRRDVSCPIRQVPTGKKQVPLIPDLNQTGKGNPSLAVSSNEFEPNSHWKSY 248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 193 NDVC-----ANDXC-----QIPF-----NLKIRANELD-----VLK-- 218
OY 249 SLLEFVTRP-----GRPEFGMANGETNENIDVNEELPARRRNRDEGEKTYVAQMT 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 219 KLVFGRPRPLDNKDNVGMEDYIKNNKTIENTI--NELI-----ESKRTIDKNN 268
OY 301 VEDKRRRLQLDGEYVAM--OEMEC-----PISKRAWTETILDG---KRL 343
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 269 ATKEEKKKLYAOADLSYKQLEBANHLISYLEKRIDTLKKNENIKELKINEIKNP 328
OY 344 PPEFESOGPTLOFTLRWTGETNDKSTAPLAPLRATRNSES-----LHOENKPG 392
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 329 PRANS--GNTPNLTLDKNNKKEIEKEIKETIAKTF--NIDSLFTDPLEYLEYLREKNK-- 384
OY 393 SVKRPQTAVKESLTLDQTRKEKDPNENRQKRLRFLYQFLN--NNTQOQTEARDLHC 450
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 385 -----NIDISAKVET-----KESTEPNTEPN--GVTYPLSYNDINNALNELNSFGDLIN 431
OY 451 PW-----CTLNCRKLY--SLTKHLKCHSRETFN--YUHPKARI--DVS----- 490
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 432 PPDYKKEPSKNLYTINERKKEFINEIKERKIKIEKKIESDKSEYEDSKSLANDITKYEKL 491
OY 491 INECYDGTAGNPQ--DIHRQPGAFASRNGPVKRPPIYHILVCRPKATYASNEF--LE 545
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 492 LNEIIDSCKENNIDLTNFEKMGKRRYS-----YKVEKLTH-----HNTFASYENSKHNE 541
OY 546 SDDGEVEQORTYSQGH---NRLYFHSOTCLPLRQEMEVDSED--EKDEWMLREKTIYQI 600
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 542 KLTAKLKTWEDYSLRNIVVEKELKYKNLISKIE--NEITLVENIKKDEQLFEKKITK- 599
OY 601 EEFSDVNEGEKEVAKLMLN--HYMKHGFADNOMNHACMLFVENYGOKIIRKKNCRFMFLH 659
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 600 ----DENPKDEKILFESDLYKVQYQVQLMKNKID-----ELKQTLKLNKVELKHNH 648
OY 660 LVSMH-----DENLISTMSIDKAVTKIREMOQKLE-----KESASPAANEIT 702
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 649 VNSYKOEKOEPPYLI---VLKKEIDKLTVPMPKVESLINEEKNIKTEGQSDNSEPST 705
OY 703 EEOGNGANGFSEINKSEK---ALETDSVSGVSGKOSK 736
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 706 E---GEITGQATTKPQQAQAGSALBDSYQAQAQOEKO 739

```

RESULT 12

US-10-108-605-303
Sequence 303, Application US/10108605
Publication No. US20020160934A1
GENERAL INFORMATION:
APPLICANT: Broadus, Julie
APPLICANT: Stamm, Lynn
APPLICANT: Bachmann, Jane
APPLICANT: Kandar, Kim

TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCOD
FILE REFERENCE: 31133B
CURRENT APPLICATION NUMBER: US/10/108, 605
CURRENT FILING DATE: 2002-03-27
PRIOR APPLICATION NUMBER: US 09/761,142
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/176,418
PRIOR FILING DATE: 2000-01-14
NUMBER OF SEQ ID NOS: 361
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 303
LENGTH: 1569
TYPE: PR
ORGANISM: Drosophila melanogaster
US-10-108-605-303

Query Match 3.7%; Score 144.5; DB 14; Length 1569;
Best Local Similarity 19.5%; Pred. No. 0.022;
Matches 104; Conservative 69; Mismatches 188; Indels 173; Gaps 22;

```

OY 9 GGGGSGSPAGSGGCG-----FGGSAVAATAATAGSGGSGCGGSGGSY 52
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 849 GGVGGGSGSRSGDAKRRYIRKKERNNSNKEAGSAAVAAGAGGNGSSSYGGD 908
OY 53 SASSSSSAAAGAAVLPYKKPRMEHY-----QADHEFLQAEPKPTQIYRFLRTNLA 107
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 909 SONGTGTATTS--SSTPTTTLTVANTARGKQNPQ---TGANSSSLMPSRQO--- 960
OY 108 PIFLHRTLTYSHRNSRTNIK-----RKTFFVDMLSKEVKKEGEOESHSL---SAHQL 159
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 961 -----QRRHSPSTSIASLCAKARKSATNSSNSYQOQONHHGTRKINNHL-- 1010
OY 160 TETGE-----FKNDKSPSPSENGNS-----VTEVLTVVCHKKRRDVSC 201
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1011 --NGTGDQGEOSRKRNSPSSNKHNETSERRVRPIYDINNVIYPSMLQSKMEI-I 1067
OY 202 PIROYV-----TGKROYPLIPDLNQTTP 224
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1068 PYKEIPIPKRIYDSNDKSHSDSMERKLSNGCVASKSEELAKOQLPPPDQOQEQP 1127
OY 225 GNP-----SLAVSSNEFEPNSHMYKSYSLFRTVRRPGRRPENGINGETNENIY 276
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1128 ALQPPRVNGLKEQAQVKNNNNNNTNNNNK-----NGLVNNAK-- 1167
OY 277 NEELPARRRNRNDEGEKTYVAQMTVPDKRRRLQLDGEYVAMOEMECPISKRATWET 336
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1168 -DELKAAEDHDQSKEDDDPKKVVYK-----LNG-----NLKNPIDKTAETQK 1213
OY 337 ILDKRLPPEFESOGPTLOFTLRWTGETNDKSTAP--IAPPLATRNSESILHOENKPGS 393
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1214 IQAPAREPP--LPKRPKLE-----TPTSNETYKANGQLVQOQLESRESEDEPQH----- 1260
OY 394 VAPRTQIYAVKESLTLDQTRKEKDPNENRQKRLRFLYQFLN--NNTQOQTEARD 447
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1261 -----VDELDSDAFTMRQALLERRRRFTLKPW--STRSRANRRVD 1304

```

RESULT 13

US-09-736-457-1812
Sequence 1812, Application US/09736457
Patent No. US20020168637A1
GENERAL INFORMATION:
APPLICANT: Wang, Tonglong
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedvick, Tom
APPLICANT: Carter, Darlick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
APPLICANT: Fan, Liqun
APPLICANT: Wang, Aijun

```

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1812
; LENGTH: 474
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-736-457-1812

Query Match      3.7%; Score 142.5; DB 10; Length 474;
Best Local Similarity 39.0%; Pred. No. 0.0055;
Matches 41; Conservative 10; Mismatches 25; Indels 29; Gaps 5;

OY      6 HGGGGGSGSPSAGSGGFGGSA-----AVAAATASGKS-----GGGS 45
Db      166 HGGGGGSGSSNAGGGGAGGANSKPAOKKSCGSKVAGAGGVSKPRAKLILAGG- 224

OY      46 CGGGGYSASSSSSAAAG-AAVLPVKKPKMEHVQADHLEFLQA 89
Db      225 -GGGKAAAAAASFPAEQAGAAALPL-----GAADHHSIYKA 262

RESULT 14
US-09-902-941-1812
; Sequence 1812, Application US/09902941
; Patent No. US20020172952A1
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tonglong
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Marnerakis, Margarita
; APPLICANT: Carter, Derrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvik, Thomas S.
; APPLICANT: Bangur, Chaltanya S.
; APPLICANT: McNabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.478C17
; CURRENT APPLICATION NUMBER: US/09/902,941
; CURRENT FILING DATE: 2001-07-10
; NUMBER OF SEQ ID NOS: 2002
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1812
; LENGTH: 474
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-902-941-1812

Query Match      3.7%; Score 142.5; DB 10; Length 474;
Best Local Similarity 39.0%; Pred. No. 0.0055;
Matches 41; Conservative 10; Mismatches 25; Indels 29; Gaps 5;

OY      6 HGGGGGSGSPSAGSGGFGGSA-----AVAAATASGKS-----GGGS 45
Db      166 HGGGGGSGSSNAGGGGAGGANSKPAOKKSCGSKVAGAGGVSKPRAKLILAGG- 224

OY      46 CGGGGYSASSSSSAAAG-AAVLPVKKPKMEHVQADHLEFLQA 89
Db      225 -GGGKAAAAAASFPAEQAGAAALPL-----GAADHHSIYKA 262

RESULT 15
US-09-849-626-1812
; Sequence 1812, Application US/09849626
; Publication No. US20020197669A1
; GENERAL INFORMATION:
```

```

; APPLICANT: Bangur, Chaltanya
; APPLICANT: Fanger, Gary
; APPLICANT: Wang, Aijun
; APPLICANT: Wang, Tonglong
; APPLICANT: Switzer, Anne
; APPLICANT: McNeill, Patricia
; APPLICANT: Clapper, Jonathan
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C16
; CURRENT APPLICATION NUMBER: US/09/849,626
; CURRENT FILING DATE: 2001-05-03
; NUMBER OF SEQ ID NOS: 1926
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1812
; LENGTH: 474
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-849-626-1812

Query Match      3.7%; Score 142.5; DB 10; Length 474;
Best Local Similarity 39.0%; Pred. No. 0.0055;
Matches 41; Conservative 10; Mismatches 25; Indels 29; Gaps 5;

OY      6 HGGGGGSGSPSAGSGGFGGSA-----AVAAATASGKS-----GGGS 45
Db      166 HGGGGGSGSSNAGGGGAGGANSKPAOKKSCGSKVAGAGGVSKPRAKLILAGG- 224

OY      46 CGGGGYSASSSSSAAAG-AAVLPVKKPKMEHVQADHLEFLQA 89
Db      225 -GGGKAAAAAASFPAEQAGAAALPL-----GAADHHSIYKA 262
```

Search completed: September 2, 2003, 11:18:12
Job time : 61.6986 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 2, 2003, 11:01:37; Search time 24.834 Seconds
(without alignments)
2861.747 Million cell updates/sec

Title: US-09-874-162a-5

Perfect score: 3885
Sequence: 1 MAPKHGGGGGGSGSPSAGS.....KALETDSVSGVSKQSKOKL 739

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: 1: PIR_76:.*
2: PIR1:.*
3: PIR2:.*
4: PIR3:.*
5: PIR4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	168.5	4.3	3968 2	A44265
2	157	4.0	622 2	I37984
3	154.5	4.0	1289 2	B84505
4	153.5	4.0	1711 1	A47392
5	153	3.9	549 2	B32372
6	153	3.9	581 1	KRMS2
7	152	3.9	427 2	A32372
8	151.5	3.9	393 2	T20268
9	151	3.9	464 2	A56600
10	149.5	3.8	2639 2	T31328
11	148.5	3.8	643 1	KRHU2
12	148.5	3.8	1611 2	T06677
13	148	3.8	863 2	T51002
14	146.5	3.8	1235 1	PMBYH
15	146.5	3.8	1639 2	S05603
16	145	3.7	1744 2	JH0720
17	144.5	3.7	1640 2	A24594
18	144.5	3.7	1875 2	S38173
19	144	3.7	592 2	E82759
20	144	3.7	594 2	B86456
21	143.5	3.7	420 2	A37343
22	143	3.7	270 2	A60830
23	143	3.7	385 2	T20410
24	142.5	3.7	474 2	I38240
25	142.5	3.7	1226 2	T24045
26	142.5	3.7	2038 2	A43742
27	142	3.6	1119 2	B70126
28	141.5	3.6	255 2	B84777
29	141.5	3.6	291 1	S31415

30	141.5	3.6	582 2	E84771	fertilization-indu
31	141.5	3.6	1087 2	T30844	serine-repeat anti
32	140	3.6	338 1	KNMU	glycine-rich cell
33	140	3.6	569 1	KRMSE1	keratin, 59K type
34	140	3.6	570 2	S07330	keratin, epidermal
35	139.5	3.6	912 2	B44450	ubiquitin-specific
36	139.5	3.6	1085 2	S62516	hypothetical colle
37	139	3.6	171 2	H84709	probable glycine-r
38	139	3.6	252 1	S01821	glycine-rich prote
39	139	3.6	419 2	A25438	keratin, type I cy
40	139	3.6	481 2	A35628	loricrin - mouse
41	139	3.6	1585 2	T31611	hypothetical prote
42	139	3.6	2500 1	WMH021	HIV-EP2 enhancer-b
43	138	3.6	1701 2	A26858	major merocrine su
44	138	3.6	5170 2	T15348	hypothetical prote
45	137.5	3.5	396 2	T49109	glycine-rich prote

ALIGNMENTS

RESULT 1

A44265 trithorax homolog HTX, version 2 - human

N:Alternate names: acute lymphoblastic leukemia gene 1 protein (ALL-1)

C:Species: Homo sapiens (man)

C>Date: 30-Apr-1993 #sequence, revision 18-Nov-1994 #text_change 01-Dec-2000

A:Accession: A44265; A44264; I58112; I37165; I38485

R:Kachuk, D.C.; Kohler, S.; Cleary, M.L.

Cell 71, 691-700, 1992

A>Title: Involvement of a homolog of Drosophila trithorax by 11q23 chromosomal translocation

A:Reference number: A44265; M0ID:93046667; PMID:1423624

A:Accession: A44265

A>Status: nucleic acid sequence not shown; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-396 <TKA>

A:Cross-references: GB:L04284; NID:q184393; PIDN:AAA58669.1; PID:q184394

A>Note: sequence extracted from NCBI backbone (NCBI:117729)

R:Gu, Y.; Nakamura, T.; Alder, H.; Prasad, R.; Canaanl, O.; Cmlno, G.; Croce, C.M.; Cell 71, 701-708, 1992

A>Title: The t(4;11) chromosome translocation of human acute leukemias fuses the ALL-1

A:Reference number: A44264; M0ID:93046668; PMID:1423625

A:Accession: A44264

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 63-316, 'GLINSELEK', 327, 'Q', 329, 'VR', 332, 'DKEGTPP', 340, 'T', 342, 'EDKTVVQAS 546, 'IQIESTSP', 2555-3554, 'N', 3556-3594, 'V', 3596-3899, 'A', 3901-3968 <GUL>

A:Cross-references: GB:L04731; NID:q339921

A>Note: sequence extracted from NCBI backbone (NCBI:117729)

R:Djabali, M.; Selleri, L.; Parry, P.; Bower, M.; Young, B.D.; Evans, G.A.

Nature Genet. 2, 113-118, 1992

A>Title: A trithorax-like gene is interrupted by chromosome 11q23 translocations in a

A:Reference number: I58112; M0ID:93265134; PMID:1303359

A:Accession: I58112

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1317-1700, 'DD', 1703-1936, 'H', 1938-2180, 'S', 2182-2328 <DJAB>

A:Cross-references: GB:L01986; NID:q307522; PIDN:AAA92511.1; PID:9553800

A>Note: submitted to the EMBL/GenBank/DBJ databases by R. Marchalek, 20 December 19

R:Marchalek, R.; Grell, J.; Lochner, K.; Nilson, I.; Slegler, G.; Zwickronner, I.;

Br. J. Haematol. 90, 308-320, 1995

A>Title: Molecular analysis of the chromosomal breakpoint and fusion transcripts in t

A:Reference number: I37165; M0ID:95315013; PMID:7794749

A:Accession: I37165

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1212-1603, 'GTE' <NAB>

A:Cross-references: EMBL:X83604; NID:q897757; PIDN:CAA58584.1; PID:q899268

A>Note: submitted to the EMBL/GenBank/DBJ databases by R. Marchalek, 20 December 19

R:Gu, Y.; Alder, H.; Nakamura, T.; Schleiman, S.A.; Prasad, R.; Canaanl, O.; Salto, H

Cancer Res. 54, 2327-2330, 1994

A>Title: Sequence analysis of the breakpoint cluster region in the ALL-1 gene involve

A:Reference number: I38485

A:Accession: I38485

A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1251-1486, 'G', 1488-1538 <RES>
 A:Cross-references: EMBL:004737; NID:9451554; PID:9451555
 C:Genetics:
 A:Gene: GDB:MLT; HTX; ALL-1; HRX
 A:Cross-references: GDB:128819; OMIM:159555
 A:Map position: 11q23-11q23
 A:Introns: 1338/1; 1362/3; 1406/3; 1444/3; 1493/3; 1525/3; 1566/1
 A>Note: the list of introns is incomplete
 C:Keywords: acute lymphoblastic leukemia; proto-oncogene; zinc finger
 F:1434-1456/Region: zinc finger CCHC motif
 F:1479-1506/Region: zinc finger CCHC motif
 F:1527-1556/Region: zinc finger CCHC motif
 F:1569-1596/Region: zinc finger CCHC motif
 F:1873-1900/Region: zinc finger CCHC motif
 F:1933-1955/Region: zinc finger CCHC motif

Query Match 4.38; Score 168.5; DB 2; Length 3968;
 Best Local Similarity 20.5; Pred. No. 0.098;
 Matches 180; Conservative 97; Mismatches 312; Indels 287; Gaps 39;

3 POKHGGGGGGG-----SGPSAGSGGGGGS-AAVAATASGKSGG 43
 13 PGTTGGGGGGGGRGLGAPRQRPVALLPPGPGGCGCAPSPVAVAAAAAGSSGA 72
 44 GSGGGGGSYASSSSSAAAAAAGAAVLPYKPKKEHVQADHELFLQAFKPTQYRFLR 103
 73 GVGPGAAAAAASASSSSSSSSSSSSGPALLRVPFDALQ----- 117
 104 NLTAFLPLHRTLYWMSRNRNTRKTFVVDMLSVKEMKGRHSHSLAHQLFTG 163
 118 -VSAAL-----GTNLR--FPAVFGESGGGSGGDEO-----FLG 150
 164 FFRHNDKPSNSENQSVTLVLLVYKCHKRRDVCPIROYETGKQVPLPDLNQT 223
 151 F-----GSDBE-----VRV-----RSPTRSPVKTSPKRGR 178
 224 P--GNFSLAVSNE--FEESNSHMVSYSLFRVTPGRREFNGMGETENIDVNEE 279
 179 PRGSDNSAAILSDPSYFSPPLNSETS-----GDKIKKDS 215
 280 LPARRKRRDEKTEFVAQMTVPDKNRRL-QLDGEYAVAMQEMECR-----ISK 329
 216 KSTIEKKR--PPTFGVAKIKTHGDISLPKGNKEDSLKTKRPSATFOQATIKK 272
 330 KRATWETILDGRLPPEFESQGTLOFTLRWGETNDKSTAPIAKPLATRNSESL- 385
 273 LRA-----GLSPLKSKFKETG-KIQIGRKGVQIVRRGRPPTERIKTPSVSLILNW 324
 386 HOEKPPSVKPTQIAVKESLTDLOTRKEKDPNENKRLRIF-----YQFLYNNTR 439
 325 KSPKRSKTR-----KEHLHLOKRIQLSDKALEGLSGLGLLQKGMQPLNSYR 377
 440 -QGTADDDLHCPWCTLNCKRLVSLKHLK-----LCSHRFIFYVYVHP- 482
 378 GQKGAQKKEKEMAOQGRKVKYKVINQFIIMPVSAISSRIKTPRPFIEEDYDPP 437
 483 -KGARIDVINECYDGSYAGNPQ-----DIHR--OPGFAFSRNG----- 518
 438 IKIARLSTPNRSAPSCSSSEKSSAASHSSQMSDSSSPVDTSTDSQASSEIQ 497
 519 -----PVKRPITHILVCPKPKTKASKSEF-----LSEDEGEVE 552
 498 VLPEERSSTPEVHPPLPISQSPENESNDRRSRYSERFSGRTTKLSTLAS-AP 553
 553 QOQRTYSGHRLYHSDTCLPLRQEMVEDSEKDEKEMLEKTIQIEEP-----SDVN 607
 554 QOQRTYSSPPPLTPPP---PLQD---ASSISDHP-WLMPPTPLASPLPASTPMQ 605
 608 EGEKEVAK-----LWNLHVKKHGFADQNMHACMLFVENYGOK-IINKNLCKRNFMLTVS 662
 606 GKRSILRPTFRMT--SLKHSRSEPPYFSSA-----KYAKBGLIKRPFIDNRPPPLT 657

OY 663 MDPENLISMSIDKATYTKL-----EMQO-----KLEKGSASPAHEITE 703
 DB 658 PEVYGFASGFSASGTASARLFSPLHSGTFEDMKRSPLRLARPRFPSEAHSRIFESVTL 717
 OY 704 EONGTANGFSEINSKEKALETDSVSGSKSKOKL 739
 DB 718 PSNRTSAGTSS-----SGVSNKRRKKV 740

RESULT 2
 137984
 keratin 9, type I, cytoskeletal - human
 N:Alternate names: cytokeratin 9; scatter protein 60K chain, placental
 C:Species: Homo sapiens (hmn)
 C:Date: 02-Jul-1996 #sequence, revision 02-Jul-1996 #text, change 29-Oct-1999
 C:Accession: I37984; S40307; S77921; S41161; B35449; I37943
 R:Reis, A.; Hennies, H.C.; Langbein, L.; Digweed, M.; Mischke, D.; Drechsler, M.; Sch
 Nature Genet. 6, 174-179, 1994
 A:Title: Keratin 9 gene mutations in epidermolytic palmoplantar keratoderma (EPPK).
 A:Reference number: I37984; MUID:94214498; PMID:7512862
 A:Accession: I37984
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-622 <RES>
 A:Cross-references: EMBL:X75015; NID:9453154; PID:CAA52924.1; PID:9453155
 R:Langbein, L.; Heid, H.W.; Moill, I.; Franke, W.W.
 Differentiation 55, 57-71, 1993
 A:Title: Molecular characterization of the body site-specific human epidermal cytoke
 A:Accession number: I37943; MUID:94131202; PMID:7507869
 A:Reference number: I37943
 A:Molecule type: mRNA
 A:Residues: 1-622 <LAN>
 A:Cross-references: EMBL:Z29074; NID:9435475; PID:CAA82315.1; PID:9435476
 A:Accession: S77921
 A:Molecule type: protein
 A:Residues: 28-53;62-104;167-188;199-233;241-249;295-362;374-430;450-480;579-604 <LAF
 R:Langbein, L.
 Submitted to the EMBL data library, December 1993
 A:Accession number: S41161
 A:Reference number: S41161
 A:Accession: S41161
 A:Molecule type: mRNA
 A:Residues: 1-11, 'SR', 13-622 <LAN>
 A:Cross-references: EMBL:Z29074; NID:9435475; PID:CAA82315.1; PID:9435476
 R:Rosen, E.M.; Meromsky, L.; Romero, R.; Setter, E.; Goldberg, I.
 Biochem. Biophys. Res. Commun. 168, 1082-1088, 1990
 A:Title: Human placenta contains an epithelial scatter protein.
 A:Reference number: A35494; MUID:90267446; PMID:2140676
 A:Accession: B35494
 A:Molecule type: protein
 A:Residues: 'X', 450-465 <ROS>
 C:Genetics:
 A:Gene: GDB:KRT9; EPPK
 A:Cross-references: GDB:303970; OMIM:144200
 A:Map position: 17q12-17q21
 A:Introns: 213/3; 241/2; 293/3; 347/3; 389/3; 464/2
 A>Note: defects in this gene may cause epidermolytic palmoplantar keratoderma
 C:Superfamily: cytoskeletal keratin
 C:Keywords: coiled coil; intermediate filament
 F:1-153/Domain: head #status predicted <HEA>
 F:154-459/Domain: helical rod #status predicted <ROD>
 F:460-622/Domain: tail #status predicted <TAI>

Query Match 4.08; Score 157; DB 2; Length 622;
 Best Local Similarity 20.9; Pred. No. 0.042;
 Matches 89; Conservative 63; Mismatches 165; Indels 108; Gaps 19;

OY 9 GGGGSGPSAGSGGGGGSAAVAATAAGSKS-----GGSGCG-GGGSISA----- 54
 DB 73 GGGGSGFSASLGLGGGSGRFGASGGGSGGFGGFGGSGGGGSGGSGFGG 132
 OY 55 -SSSSAAAAAGAAVLPYKPKKEHVQADHELFL--QAPEK----- 92

Db 133 LGGFGGAGGDDGILITANEKSTMOELNSRLASIDYKVALDEANNDLENKIQDWTDKG 192
 QY PTOIVR-----FLRT-----RNLAPILFHLRTLYMSHRNSRTNIK--RKTQYDDMLSK-- 140
 Db 193 PAIQKNSPYNTLIDKQJVDLTVGNKKTLIDIDNRMFLDPRIFEFEMEQNRQGV 252
 QY 141 -----VEKKGEQESHSSAHILQTLFTGFHKNDRKPSNSENEDNSVYLE 185
 Db 253 DADINGLROVLNMLMERSDLEMOYETLOEELMA-----LKNHKEEMQLGQNS---- 303
 QY 186 VLLVAVCHKKRKDVSCPIRQVPTGKKQVPLIPDLNOKPCNPPLSAVSNSEPESSAHV 245
 Db 304 -----GDVAVELNVP--GKDLTKTLNDRQ-----EYEQ--TAKNRKDIENQYET 346
 QY 246 KSYSLFLVTRPR-----REFNGINGETNENIDVNEEL--PARKNRNDEGEKTEVA 297
 Db 347 QITQLEHVESSSGQVQSAKQVTLRHQVQLEIEFLOLSKKALERSLEDGKTKRQYCG 406
 QY 298 QMTVVDK---NRRLDLDGEYEVAMQEME-----ECPISKRAETITLDGKRLPRE 347
 Db 407 QLOMQEOISNLEAQITDVRQIECONQYSLLSIKMLKEIEITYHMLBQGO--EDPE 465
 QY 348 TFSOG 352
 Db 466 SSGAG 470

RESULT 3

184505
 calcium-dependent actin-binding protein - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 02-Aug-2002
 C:Accession: I84505
 R:Walent, J.H.; Porter, B.W.; Martin, T.F.
 Cell 70, 765-775, 1992
 A:Title: A novel 145 kd brain cytosolic protein reconstitutes Ca(2+)-regulated secretion
 A:Reference number: 139186; PMID:92286596; PMID:1516133
 A:Accession: I84505
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1289 <RES>
 A:Cross-references: EMBL:U16802; NID:g577427; PIDN:AB88635.1; PID:g577428
 C:Genetics:
 A:Gene: CAPS
 C:Superfamily: calcium-dependent actin-binding protein; pleckstrin repeat homology
 C:Keywords: actin binding

Query Match 4.0%; Score 154.5; DB 2; Length 1289;

Best Local Similarity 19.1%; Pred. No. 0.16; Mismatches 283; Indels 251; Gaps 43;

Matches 157; Conservative 130; Mismatches 283; Indels 251; Gaps 43;
 QY 7 GGGGGGSGPSAGSGGCGFSGSAVAATAAGGKSGG---GSGCGGGSYSASSSSAAAA 63
 Db 44 GSAGMGSGAGAGVAGGSGGSGA-----SSGCGAGLQPPSSAGGGRSSPSPS----- 93
 QY 64 AGAAVLPAKKPKMEHVQADHLEFLQAEKPTQYRFLTRNLIAPIFLHRTLYMSHRNS 123
 Db 94 ---VSEKEKEERLEKKEE---ERKKRLQLY-----VFVHMCIAVPPNAQ 135
 QY 124 RINIRKRTKVDMLSKVEK-----MKGEQESHSSAHILQ--TTGPFHNDKQSPN- 174
 Db 136 PTDMARPRKISKQQLQTVKDRQAFNLNGETQIVADEAPMNAVQSYEYFLKSDRAARMV 195
 QY 175 -----SENQNSVTLFVLLVVKCHKKRKDVSCPIRQVPTGKKQVPLIPDLNOKPCNP 229
 Db 196 QSGGGSANDSREYF-----KHHI-----EKVRSLEPEIDGSKETVLS 233
 QY 230 L-----AVSSNEFEF-----SNSHVKSYSLLFVTRP--GRREFNGMT--NGE 269
 Db 234 SMMAKDAIYRGEDPRKQOARMTASASELLSLKQLVEMFQNIIGIKFEEHQLYNAC 293
 QY 270 TRENID-----VNEELPAR--RKRNRDEGETFYAQMNTVDKNNRLQDLGEYEVAMQEME 323

Db 294 QLDNPEQQAQIRREIDGRLQAMADQIARERKFPKFKVSKEMENNTIEELKSVNLMANIE 353
 QY 324 ECPISK-----KRAETITLDGKRLPRETFESQPTLOFTLRMT--GETNDKSTAP 372
 Db 354 SMPVSGGGEFKLQKLRNSNASTIDMGSEENDLSQVLSLEVLSTLEVIMEOGLKSLAP 413
 QY 373 ---IAKPLATRNSESILH-----QENKP-----GSVKEPTQTI-AVRESL---TTDLQTRKER 416
 Db 414 NRIYCTMEVEGGERLQTDQAEASKPTMGQGFSTTHALPAVKVLFTESTGVALLERK 473
 QY 417 D-----TPNNRQKLIFFQFLYNNNTROQTARDDLHCWCLNKRKLYSLKHLK 468
 Db 474 ELGRVILHPTNPSPKO-----SEHKKKTYSK---NCPDQDLKIKLAVRMDKPN 519
 QY 469 LCHSRFIFN-----YUHHKGARIDVINECYDGSAGNPQDINHROPGFASHNGPVKRT 523
 Db 520 MKHSGYLMTIGKNVWRMKRRKRFVLAQVQ-----YTFV----- 553
 QY 524 PITHILVCRPKRTKASKSEFLSEDEGV---EQRTYSSGHNFLYHS---DTCLPLRP 576
 Db 554 -----MCSYREKKAQPOEILQIDGYTVYTDPOPLEGG--RAFFNAVEGDTVI---- 601
 QY 577 QEMEVSEDEKDP--EWLR-----EKTI---TOLEFSDVNEG-----EKVMTKLMLNH 620
 Db 602 ---FASDDEQDRIILVQAMTRATGOSHRFPPTQVOKLKPRAETCLSMAPISQFYADR 657
 QY 621 VNRHG---FIADQNQN--HACMLFV-----ENY-----GQRIIRKNLC--RN 655
 Db 658 AQKHGMDERTSSRPNCFDASLIFEMQORLTLDRLNDLSQVLPFGQVFLVDEYCARN 717
 QY 656 FMLHLVSMDFNLISMSIDKAVTKLREMOQKLEKESASP 696
 Db 718 -----GVRGCHRHLCYRDLLEALRANGAMIDP 744

RESULT 4

A47392

chromodomain-helicase-DNA-binding protein, CHD-1 - mouse

N:Alternate names: KYBP protein

C:Species: Mus musculus (house mouse)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Dec-2002

C:Accession: A47392; S21568

R:Delmas, V.; Stokes, D.G.; Perry, R.P.

Proc. Natl. Acad. Sci. U.S.A. 90, 2414-2418, 1993

A:Title: A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SWI

A:Reference number: A47392; PMID:93211972; PMID:8460153

A:Accession: A47392

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1711

A:Experimental source: S194 plasmacytoma cells

A:Note: sequence inconsistent with the nucleotide translation

R:Delmas, V.; Perry, R.P. Submitted to the EMBL Data Library, May 1992

A:Description: KYBP, a mammalian protein that contains the SNF2/SWI2 helicase domain

A:Reference number: S21568

A:Accession: S21568

A:Molecule type: mRNA

A:Residues: 772-1711 <DE2>

A:Cross-references: EMBL:X66028

C:Superfamily: chromodomain helicase CHD1; chromobox homology

C:Keywords: DNA binding

F:293-336/Domain: chromobox homology <CB1>

F:387-427/Domain: chromobox homology <CB2>

Query Match 4.0%; Score 153.5; DB 1; Length 1711;

Best Local Similarity 20.1%; Pred. No. 0.27; Mismatches 124; Conservative 80; Mismatches 207; Indels 205; Gaps 31;

QY 5 KHGGGGGSGPSAGSGGCGFSGSAVAATAAGKSGGSGCGGGSYSASSSSAAAAA 64
 Db 11 RNSGSSGSGGDCGSGSAGSGSGSSSSSDSSSOS--GSSDSDSGSDSGSSESSESDTS 69

RESULT 7

A32372

female-specific doublesex protein - fruit fly (*Drosophila melanogaster*)C:Species: *Drosophila melanogaster*

C:Date: 02-Nov-1989 #sequence_revision 02-Nov-1989 #text_change 23-Feb-1997

C:Accession: A32372

R:Butler, K.C.; Baker, B.S.

Cell 56, 997-1010, 1989

A:Title: *Drosophila* doublesex gene controls somatic sexual differentiation by producing

A:Reference number: A32372; MUID:89168451; PMID:2493994

A:Accession: A32372

A:Status: preliminary

A:Molecule type: DNA, mRNA

A:Residues: 1-427 <BUR>

A:Cross-references: GB:M25292; GB:J03156

C:Genetics:

A:Gene: FlyBase:dax

A:Cross-references: FlyBase:FBgn0000504

C:Keywords: alternative splicing; DNA binding; nucleus; transcription regulation

Query Match 3.9%; Score 152; DB 2; Length 427;

Best Local Similarity 27.3%; Pred. No. 0.052;

Matches 45; Conservative 25; Mismatches 47; Indels 48; Gaps 4;

OY 11 GGGGSPASAGGGCGFGSAVAATATAGKSGGCGGGGYSYSA----- 54

DB 242 GGSVGPATSSSGGAPSSSNAATAATSSNGSGGGGSGGAGGGRSGTSVITSA 301

OY 55 -----SSSSSAAAAAGAAVLP-----YKPKMEHQADHFLQAE 91

DB 302 DHMTTVPPTPAOSLEGSCDSSSPSSSTGAILITISVYNKNGAMPPLGODVPLDYCQ 361

OY 92 KPTQYRFLRNLRLAPILFHLHTLYMSHRNSTRNKRKTFKVD 136

DB 362 KLEKFRX-----PMEL-MPLMYVILKQADANIEASRIEE 397

RESULT 8

T20268

hypothetical protein C56A3.1 - *Caenorhabditis elegans*C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T20268

R:Slims, M.

submitted to the EMBL Data Library, July 1996

A:Reference number: Z19244

A:Accession: T20268

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-393 <WILL>

A:Cross-references: EMBL:Z77655; PIDN:CA01137.1; GSPDB:GN00023; CESP:C56A3.1

A:Experimental source: clone C56A3

C:Genetics:

A:Gene: CESP:C56A3.1

A:Map position: 5

A:Introns: 51/3; 91/1; 121/1; 331/3

Query Match 3.9%; Score 151.5; DB 2; Length 393;

Best Local Similarity 34.6%; Pred. No. 0.05;

Matches 46; Conservative 8; Mismatches 46; Indels 33; Gaps 5;

OY 7 GGGGGGSPASAGSGGF--GSAVAATA-----TASGKSGG 43

DB 117 GGGGGGGGGGGGGGGGGGASGSGGASAPVSLPAPSYGPPPPAPSPSHAPSGGYSYG 176

OY 44 GSCGGGGYSASSS-----SSAAAAAGAAVLPYKRRKMEVQADHFLQAEKRP-----TQ 95

DB 177 GSSGGGYSGGGGGGGAGAAAGATAAAGV--DEASGSGEPPVQVDFHSPKPECTQ 234

OY 96 YRFLRNLRLAP 108

DB 235 QVYTIMLSRKYP 247

RESULT 9

A56600

intermediate filament protein IF-1, cytosolic - common lancelet

C:Species: *Branchiostoma lanceolatum* (common lancelet)

C:Date: 11-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 21-Jul-2000

C:Accession: A56600; S24265

R:Riemer, D.; Dodemont, H.; Weber, K.

Eur. J. Cell Biol. 58, 128-135, 1992

A:Title: Analysis of the cDNA and gene encoding a cytoplasmic intermediate filament (

family)

A:Reference number: A56600; MUID:92354564; PMID:1644059

A:Accession: A56600

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-464 <RIE>

A:Cross-references: EMBL:X64522; NID:95724; PIDN:CAA5827.1; PID:95725

A>Note: sequence extracted from NCBI backbone (NCBI:P110465)

A:Note: Intron positions were also determined

C:Superfamily: cytoskeletal keratin

C:Keywords: cytosol; intermediate filament

Query Match 3.9%; Score 151; DB 2; Length 464;

Best Local Similarity 19.8%; Pred. No. 0.067;

Matches 95; Conservative 77; Mismatches 158; Indels 150; Gaps 20;

OY 7 GGGGGGSPASAGSGGFG--GSAVAATAATAGKSGG--SCGGGGYSASSSSAAAA 63

DB 18 GGGGGGGGGRASYSSSGRRSGGGRKAGCVRSSSVSSYAGGGGRVNAAGVMA 77

OY 64 AG--AAVLPYKRRKMEVQADHFL-----FLQAEKPTQYRFLRNLRLAPILFLH 112

DB 78 VGGPAGALQTLSDARTRAHKQELSHLNDRFASYDK---VRYLQERNSKLEAQIKIQ 133

OY 113 RT-----LTYMSHNSNTNKRKFKY-----DD 136

DB 134 ESRKAPNKIDYKELRDLRALVDELNSKADLEIRNNWQOQADYKRWODEAGLSE 193

OY 137 MLKYKMKGEQESHSLAHLQTLFTGFHKNDKPSNSEPNQSVTLFLVYKCHKRR 196

DB 194 LEAEIERLKKELDAATM-ARLDL-----ENLSTQGEIDIDLRVHDE 236

OY 197 KDVSCEPIQVPTKKQVPLIPDLNQTGPNF---PSLAVSSNEFPSSNHWKSYSLPR 253

DB 237 -----IROLQDLNLSLTIIVEVDSRASTFAPGPDLTALREIR-----TOYELGR 283

OY 254 VTRPGRRNGMNGETNNIDVN-----EELPARKRRREGKTFVQAQTVFDKNR-- 306

DB 284 VNR-----EDADVKKQKPESELAHORE--DNEALMTARTEVTELRNL 325

OY 307 -----RLQDLQGEYVNAQEMEE-----CPISKRRATW 334

DB 326 NSLVAENENLKKNAHLESLAELEKRMQLEIEYQAAIRDLQLELETSTSEKQOMQV 385

OY 335 ETLIDGKRLPPRETSQGPILQ-FTLRWTGETNDKSTADIANPLARNSLSLHOKNPGS 393

DB 386 KMLMDTKALDMEIAAYRKLGEIEIRLGESEKSG-----YQOTSSSSSSSTQYKSSGS 440

RESULT 10

T31328

fibroin - Chinese oak silkworm

C:Species: *Antheraea pernyi* (Chinese oak silkworm)

C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999

C:Accession: T31328

R:Sezutsu, H.; Tamura, T.; Yuhkihiro, K.

submitted to the EMBL Data Library, August 1998

A:Description: Characterization of the full length fibroin gene of a wild silkworm, A

A:Reference number: Z20995

A:Accession: T31328

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2639 <SEZ>


```

Db      545 SSTAGSTLKKCGSGSSR---HRRSMNVFGSSSSASPRAGLSRFSFSSHRNV----- 593
Qy      168 NDRPSPNSENQNSTYLEVLLVYCHKKRKDVSCPIROVPTGKQVPLDNLQTKPGNF 227
Db      594 -----PSKERISOSSLTSE-----KILEVPLTE-GSGNKLIYKLP-----RGSS 630
Qy      228 PSLAVSSNEFE--PSNSHMYKSYSLFRVTPRPGRRFNGMNGTENNIDVNEELPAR 284
Db      631 PAQSVSGSLDEDPAPVNV-----RVSSP-----VHTKQELCDNN---WME 668
Qy      265 KRNREDGEKTFVAQTFVDKNNRLLDGEYE-----VAMQEBECPISKRAWTETI 337
Db      669 KNHSYRADVSSVYLAESWQSNELKDILTGSQEATGSPVLVAGDEREGALKDKASGNV 728
Qy      338 LDGKRLPPEET-----FSQGPLTOLFTBWTGETN-----DKS 369
Db      729 ATSSLGNEFKSGERRGCTLSSNNALIESCVRTS-ETNNSLAGSDVGMNLLASVAADENS 787
Qy      370 TAPIAKPLATRNSSELIHOEN-----KPGSVKPTOTIAYKESLTTDQ 411
Db      788 KSPVASPSVSGOPNLSMENSIVGNNTKLMASDGLPHQHOAVRPTLSNEQGEQYVSSG 847
Qy      412 TRKEKDTNENRQKLRIFYOFLYNNNTROOFEARDDLHCPCWCTLCRKLISLKHKLCH 471
Db      848 TQLESEIKNESKTDGRV-----KSSNSDTE--DL----- 874
Qy      472 SFNFINYYHPRGARIDVSYNECYDSYAGN--PODIHROGFARFNRGPKV--RPIPIH 527
Db      875 QRFV-----DQALE--SNENSDGVYASPLPKVYIKENLLDSDSEVADIKTDVNS 924
Qy      528 ILVCPKPKTKASMSFEL-ESED---GEVEQOQRTYSQGNHRLYFHSPTCLPLRQEMEV 581
Db      925 EADCTSDTLKRYVASMLTECRDVSQKWDVSAVEHT-----PL--EGYVD 966
Qy      562 DSEDEKDEMLREKITTQIEESDYNEG-----EKRYKLMNLHYMKHGFADNQ 631
Db      967 DKREKRPALSELVYKVEEDVPVSSGISRCMDAVSIDRPTTEVNV-----NTAFNH 1019
Qy      632 NNHACMLFEVNYGQKLIKNNLCRNFMILHLYSHMDFNLISIMSDKAVT--KIREMOQKLE 689
Db      1020 MD-----QADIKK-IKQDPDTYSGAVK----ASAGLDSVSYGKVEPVGNI 1063
Qy      690 KGESASPAHEITEBONG 707
Db      1064 -----NIEMERYSG 1073

RESULT 13
hypothetical protein B7J19.10 [Imported] - Neurospora crassa
C:Species: Neurospora crassa
C>Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C:Accession: T51002
R:Schulte, U.; Aign, V.; Hohnsels, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, July 2000
A:Reference number: 225286
A:Accession: T51002
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-863 <SCH>
A:Cross-references: EMBL:AL389899; GSPDB:GN00116; NCSP:B7J19.10
A:Experimental source: BAC clone B7J19; strain OR74A
C:Genetics:
A:Gene: NCSP:B7J19.10
A:Map position: 6
A:Introns: 782/3

Query Match      3.8%; Score 148; DB 2; Length 863;
Best Local Similarity 18.0%; Pred. No. 0.24;
Matches 114; Conservative 78; Mismatches 228; Indels 212; Gaps 25;
Qy      247 SYSLLEFRVTPGRREFNGMNGTENNIDVNEELPARR-----RNREDGEKTFVAQM-TV 301

```

```

Db      199 SYSLLEIESAG-----DPMPPSSIMPAKVEDPLVRNLPROMALYQALPDI 246
Qy      302 F-DKNRRQLLDGEVEVAMQEBECP---ISKRAWTETIILGKRLPPEITSQGPITQF 357
Db      247 FHTRNKKYVRL---KARVQSQREPTPTDFVMDIVRMQVSISS----- 285
Qy      358 TLRTWGETNDKSTAPI-----AKPLA-----TRNSESL---HOENPGSVKPTQ 398
Db      286 ---QLTGNQPKELMPSITVIDPWAPKRPISLPLVNGTNCVKGINGHHDPNTSNEIRK 343
Qy      399 TIAVESLTTDQTRERKD--TPNENRQ-----KLRIFYOFLYNNNTROOFEARD--- 447
Db      344 NMEVNSTIADIEELAEGLDLPNRSRHRHTEVNVYROLMTNPAVGEKPRRRRADDHP 403
Qy      448 -----LHCPGTLNCRKLYSLNLHKLCHSRFTFNYY 480
Db      404 LDEHTVTVYLLPPTPGNEETLQATCNKLSCLCAAEHRIQLRAHFS-CHPEYEENF-- 460
Qy      481 HPRGARIDVSYNECYDSYAGNPDIHROGFA-----FSRNGPVKRPPIPIH 528
Db      461 -----DQKKGATL---VAVHAPGNTSTMPPLSKLFSGLPYKRLDLSKF 505
Qy      529 L-----VCRPKTKASMS---EFLESEDEVEQOQRTYSQGH----- 561
Db      506 VNGDDSWVKSRLLPDPNDCERPHLPASRSPTKGPQOQOQOQOQOQOQOQOQOQOQ 565
Qy      562 PTTSTRSRAAAATTKPKQKPYIPHNIRPYD-----PLKVELAPGSEVRYPPLDG 620
Db      591 WLREKTIQIEESDYNEGEKRYKLMNLHYMKHGFADNOMNHACMLFEVNYGQKLIK 650
Qy      621 WLTKRADALGESDYDEPEKEKYMGMQDAVYIIQKHLCSQYLPREFRNVREKATLLEK 680
Db      651 NL-----CRNFMILHLYSHMDFNLISIMSDKAVTKLEMOQKLEKGESASPAHEITEB 706
Qy      681 RSRABEIKHMAVLLARVYDATTVAAYVR---ELNEARKAMAAESA-----N 726
Qy      707 GFANGFSEINSKEKALETDYSVGSVSKOSKOK 738
Db      727 GPGAGDAGASKESEAEQOQOQOQOQOQSSGRKK 758

RESULT 14
PMBYH
potassium transport protein TRK1, high-affinity - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein J0693; protein YJL129c
C:Species: Saccharomyces cerevisiae
C>Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 12-Nov-1999
C:Accession: S05849; S56910
R:Gaber, R.F.; Styles, C.A.; Fink, G.R.
Mol. Cell. Biol. 8, 2848-2859, 1988
A:Title: TRK1 encodes a plasma membrane protein required for high-affinity potassium
A:Reference number: S05849; M0ID:88302204; PMID:3043197
A:Accession: S05849
A:Molecule type: DNA
A:Residues: 1-1235 <GAB>
A:Cross-references: EMBL:Z49404; NID:g1008329; PIDN:CAAB9424.1; PID:g1008330; GSPDB:G
C:Genetics:
A:Gene: SCD:TRK1; MIPS:YJL129c
A:Cross-references: SCD:S0003665; MIPS:YJL129c
A:Map position: 10L
C:Superfamily: potassium transport protein TRK1
C:Keywords: ion transporter; potassium transporter; transmembrane protein
F:50-70/Domain: transmembrane #status predicted; <TM01>

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 2, 2003, 11:01:37; Search time 23.3289 Seconds
(without alignments)
1340.299 Million cell updates/sec

Title: US-09-874-162a-5

Perfect score: 3885
Sequence: 1 MAPKRGSGGGGGSGPSAGS.....KALETDSVSGVSKSGKQKRL 739Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database:

1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/5C.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/5D.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	160	4.1	3969	3 US-08-061-376-5	Sequence 5, Appl1
2	153	3.9	738	3 US-08-864-038A-3	Sequence 3, Appl1
3	147	3.8	832	1 US-08-209-747-2	Sequence 2, Appl1
4	147	3.8	832	1 US-08-458-298-2	Sequence 2, Appl1
5	146.5	3.8	1235	1 US-08-118-101A-2	Sequence 2, Appl1
6	142.5	3.7	474	4 US-09-702-705-1812	Sequence 1812, Ap
7	142.5	3.7	474	4 US-09-736-457-1812	Sequence 1812, Ap
8	142.5	3.7	712	4 US-09-059-584-51	Sequence 51, Appl1
9	141	3.6	201	3 US-09-052-995-1	Sequence 1, Appl1
10	141	3.6	201	3 US-09-053-003-40	Sequence 40, Appl1
11	141	3.6	201	4 US-09-054-281-22	Sequence 22, Appl1
12	141	3.6	201	4 US-09-478-948-6	Sequence 6, Appl1
13	139.5	3.6	912	1 US-07-789-915A-8	Sequence 8, Appl1
14	139.5	3.6	912	1 US-08-005-002C-8	Sequence 8, Appl1
15	139.5	3.6	912	1 US-08-487-203A-8	Sequence 8, Appl1
16	139	3.6	100	4 US-09-411-067C-4	Sequence 4, Appl1
17	137.5	3.5	647	2 US-08-770-761A-8	Sequence 8, Appl1
18	137.5	3.5	705	2 US-08-770-761A-7	Sequence 7, Appl1
19	137	3.5	235	2 US-08-529-190B-1	Sequence 1, Appl1
20	137	3.5	641	4 US-09-249-585A-3	Sequence 3, Appl1
21	137	3.5	641	4 US-09-410-399-4	Sequence 4, Appl1
22	134.5	3.5	689	4 US-08-778-570B-16	Sequence 16, Appl1
23	134.5	3.5	689	4 US-09-059-584-16	Sequence 16, Appl1
24	134.5	3.5	709	4 US-08-778-570B-15	Sequence 15, Appl1
25	134.5	3.5	709	4 US-09-059-584-15	Sequence 15, Appl1
26	131	3.4	2211	4 US-09-738-884-1	Sequence 1, Appl1
27	129.5	3.3	1332	1 US-07-609-716-41	Sequence 41, Appl1

28	129.5	3.3	1332	3 US-08-475-411A-41	Sequence 41, Appl1
29	129.5	3.3	1332	3 US-08-478-029A-41	Sequence 41, Appl1
30	129.5	3.3	1494	3 US-08-755-587-186	Sequence 186, Appl1
31	129	3.3	129	4 US-09-916-109-5	Sequence 5, Appl1
32	129	3.3	483	4 US-09-916-109-4	Sequence 4, Appl1
33	127.5	3.3	146	1 US-07-609-716-105	Sequence 105, Appl1
34	127.5	3.3	146	3 US-08-475-411A-105	Sequence 105, Appl1
35	127.5	3.3	146	3 US-08-478-029A-105	Sequence 105, Appl1
36	127.5	3.3	609	4 US-09-107-532A-4451	Sequence 4451, Appl1
37	127.5	3.3	1345	2 US-08-977-767-3	Sequence 19, Appl1
38	127	3.3	651	3 US-08-556-978B-19	Sequence 19, Appl1
39	127	3.3	651	3 US-09-247-806-1	Sequence 1, Appl1
40	127	3.3	718	1 US-08-425-069-2	Sequence 2, Appl1
41	127	3.3	718	2 US-08-317-844B-2	Sequence 2, Appl1
42	127	3.3	747	3 US-09-034-177-3	Sequence 3, Appl1
43	126.5	3.3	40	3 US-08-815-190A-17	Sequence 17, Appl1
44	126	3.2	123	4 US-09-072-596-243	Sequence 243, Appl1
45	126	3.2	142	1 US-07-609-716-100	Sequence 100, Appl1

ALIGNMENTS

RESULT 1
US-08-061-376-5
Sequence 5, Application US/08061376
Patent No. 6175000
GENERAL INFORMATION:
APPLICANT: Evans, Glen A.
APPLICANT: Djabbal, Malek
APPLICANT: Seller, Licia
APPLICANT: Parry, Pauline
TITLE OF INVENTION: CHARACTERIZATION OF A CHROMOSOME 11023
TITLE OF INVENTION: TRANSLATION BREAKPOINT ASSOCIATED WITH ACUTE LEUKEMIAS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretly, Schroeder, Brueggemann & Clark
STREET: 444 South Flower Street, Suite 2000
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/061,376
FILING DATE: 13-MAY-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: P41 9387
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619)546-4737
TELEFAX: (619)546-9392
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3969 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-061-376-5
Query Match 4.1%; Score 160; DB 3; Length 3969;
Best Local Similarity 20.6%; Pred. No. 0.0008;
Matches 181; Conservative 97; Mismatches 308; Indels 292; Gaps 40;
3 POKHGGGGGG-----SGPSAGSGGGFGGS-AAVVAATSGGKSGG 43
|||||

```

Db 13 PGTGGGGGGGRLGDDPRORVALLPPGPVGGGCGCAPPSPPVAAAAAAGSGCA 72
Oy 44 GSCGGGGSYASSSSSAAAGAAVLPYKPKMEHVQADHFLFOAEKPTQIYRFLNTR 103
Db 73 GVPGAAAAAASASSSSSSSSSSSASSGALLRVGPGFADALQ----- 117
Oy 104 NLAPILPILRLTYMSHNSNTNKRKTEKVDMLSKYEKKKGGESLSAHLQLTPTG 163
Db 118 -VSAI-----GTNLR--FRAVFGESGGGGGGEDEQ-----FLG 150
Oy 164 FFHKNDKPPNSNEONSVTLEVLVYCHKKRKVDSCPIQOVTPGKQVPLIDLNQTK 223
Db 151 F-----GSDE-----VRY-----RSPTRSSVTSRKPRGR 178
Oy 224 P--GNPSLAVSNE--FEPNSHNVSYSLFLVTRPRGRREFNGMNGENIDVNEE 279
Db 179 PRSGSDRSAILSDPSVFSPLNKSETKS-----GDIKKKDS 215
Oy 280 LPARKNRREDGETFPAQMTVPFQKNRL-QLDGEYVAMQEMEBCP-----ISK 329
Db 216 KSIKKKGR--PPTFGVKIKITHGDISLPKGNKEDSLKIKRPPSATFOQATKIK 272
Oy 330 KRATWETILDGRLLPPEFSGPTLOFTLRMTGET-----NDRSTAPIAKPIATRSE 383
Db 273 LRA-----GKLPSSLSLRQSGFKZEGKCYKIZTERKASINRKDQFGL-LIILNWK 324
Oy 384 SLHQENKPGSVKPTQITAVKESLTDTLOTRKEKDPMENRQKRIE-----YOFLYNN 437
Db 325 S--PRKSGKTR-----KEHLHOKKIRQLSDKALEGLSGLGFLQKQOMPLSNS 374
Oy 438 TR--OOTFARDLHCPWCTLNCRKLYSLKHLK-----LCHSRFTFNYYH 481
Db 375 YRGOKKGAKKIEKAAQOLGKRVKQYKINRQFIMPVSAISSRIITKTPRRFIEDDYD 434
Oy 482 P--KGARIDVINECYDGYAGNPQ-----DIHR--QGFASRNG----- 518
Db 435 PRIAILESTPPNRFBAPSCGSESEKSSAASQHSOSSDSRSSSPVDISTSOASE 494
Oy 519 -----PVKRPDITHILVCRPKRTASMEF-----LESEDE 550
Db 495 IOVLEERSDTPVHPLPIISOSEPNESNDRSRYSERSFSGRTTKLSTIQS---- 550
Oy 551 VEOORTYSSGHNRLYFSDTCLPLRQEMVEDSEDEKDPERKKTITQIEEF-----SD 605
Db 551 APOQOTSSSPPLLTTPP---PLQP-----ASSISDHP--WLMPTTLPASPELPASTAP 602
Oy 606 VNEGEKEYMK-----LWNLHVAKHGFIDQNMNHACMLFVENYQOK-IKKMICHENFMLHL 660
Db 603 MCGKRKSLIREPTFRW--SLKHSRSEPOYESSA-----KYAKEGLIRKIFDNERPP 654
Oy 661 VSMHDFNLISIMSIDKAVTKLR-----EMQ-----KLEKESASPANEI 701
Db 655 LTPEDVFASGFSASGTASARLFSPLHSGTRFDMHKRSPLLRAPRFTPSAHSRIFESV 714
Oy 702 TEONGTANGSEINSKEKALEFTMSVGSYKSKOKUL 739
Db 715 TLPSNRISAGTSS-----SGVSNRKRKRKY 739

```

```

STREET: Iashinden
CITY: Tsu-city
STATE: Mie-prefecture
COUNTRY: JAPAN
ZIP: 514-01
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: Microsoft Windows 95
SOFTWARE: Word Perfect 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/864,038A
FILING DATE: May 28, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 8-184459
FILING DATE: 15-July-1996
ATTORNEY/AGENT INFORMATION:
NAME: C. Bruce Hamburg
REGISTRATION NUMBER: 22,389
REFERENCE/DOCKET NUMBER: F-5610
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)986-2340
TELEFAX: (212)953-7733
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 738
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Pinctada fucata
CELL TYPE: mantle epithelial cell
FEATURE:
NAME/KEY: peptide
LOCATION: from 1 to 738
IDENTIFICATION METHOD: E (by experiment)
US-08-864-038A-3
Query Match 3.9%; Score 153; DB 3; Length 738;
Best Local Similarity 43.5%; Pred. No. 0.0024;
Matches 37; Conservative 11; Mismatches 11; Indels 26; Gaps 3;
Oy 7 GCGGGGSGPAGSGGCGGCGGSAVVAATASGKSGC-GSCGGGGSYSAS----- 55
Db 437 GGGGGGCG--GCGGGGCGAGALAAALAAAGAGGCGGAGALAAALAAAGAGCGG 494
Oy 56 -----SSSSAAAAAGAA 67
Db 495 FCGICGLGGLGGSAAAAAANA 519

```

```

RESULT 3
US-08-209-747-2
Sequence 2, Application US/08209747
Patent No. 5733771
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.
TITLE OF INVENTION: cDNAs Encoding Minor Ampullate Spider
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-3487
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

```


MOLECULE TYPE: protein
US-08-118-101A-2

Query Match 3.8%; Score 146.5; DB 1; Length 1235;
Best Local Similarity 20.1%; Pred. No. 0.0019;
Matches 154; Conservative 90; Mismatches 227; Indels 295; Gaps 39;

```

QY 97 YRLRLNLAPILFLHRTLYMSHRNSTNIRKTFVYDMLSKV-----EKM 145
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 148 FKMRTKTLERELTARMTKRTQTQTSYPRKQATDDPQKLFGEVNVDEQSVH 207
QY 146 GEQSHSLANLQLTTFG-----FHKNDKPRSPNEQNSVLEVLVYVCHKR 196
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 208 SDQNSHDSISRDSSNNNTNHNSSGSLDDEFAEDETDNGEYOENN----- 252
QY 197 KDVSCPIROVPTGKQVLLPDLNQTGKPNFSLAVSNF--EPSNSHMVSYLLF-- 252
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 253 -----SYTYGSSSNVYADSLNOKPRPSSLRDEP 283
QY 253 -RYTRPGREFNGMINGETNENIDVNEELPARRRNRNDEGKTFVQMTVEFKNRRLQL 311
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 284 HSKRPAR-----VPSKFAKRRGRD-----ISPADMYRSTIML 318
QY 312 DGEYVAMQEMEGPI-----SKKRAETIILDGKRLPPEPETSQGPFL 355
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 319 QGKHE-ATADEGPPLVIGSPADGTRYKSNVNLKKAIG--INGNKIKIDKNESVTD 374
QY 356 QPTLRMTGETNDKSTAPIAKPLATRNSESLHOE--NKPGRSVKPTOT-----IAYKESL 406
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 375 QNSV--SSEAN--STAVS-----DESSLHTNFGNKVPSLR-INTHRNSGPAITDNA 423
QY 407 TTDLQTRKEDTPRENOKRIEYQFLYNNNR--QOTEARDDLHCPWCLNCRKLYS- 462
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 424 ETD-----KKGPGS-----IQFDITKPRKISKRVSTFDLLN-PKSSVLKRAKSK 468
QY 463 --LKHLKLCDSRIEFNVYHPRKARIDVSI-NECYGSSVGNODITHROGFAFSRNGP 519
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 469 KYLKKHNP-----PKARRIQOIKRRLSTGSTEKN-----SSNV 502
QY 520 VKRPTIHLVCRPKRTKASME-----FLESEDEGEVQOQRTYSSGHNRL 564
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 503 SDRKPI-----DMDDDDDDDNGDNNEEFADNNEG--EDEVOQSEP--- 546
QY 565 YFHSDTCLPLRPOEME-----VSEDEKDEPMLREKTIQIEFSDVNE 608
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 547 --HSDSELSKHOQOQEKHQLQONLHRYMKTGKSPDNDNSRAVPMERSRTI-DMAAKDINE 603
QY 609 -----GEKEVYKLMNLH-----VMKHG-FIADNQMNHACLFEVNTGOK 646
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 604 LARPPDFOKMYOVMKAKHRRKPRFRKRGNNKLFEGPYASDSDRPN--SNTG-- 658
QY 647 IIRKRLCRNFMHLVSMHDFMLISMSIDKAVTKLREMOQKLEKESASPANEETEEQN 706
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 659 -----NSILHYAE-----SILHHDS-----HNGSEASSDSNENIYISNG 695
QY 707 GT-----ANGSEINSKEKA-----LETDSVSGVSKOSKOKL 739
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 696 GSDHNGLNINYPYNDDEGYGLHFDYDLDPRHDLISKSGSKTYL 741

```

RESULT 6
US-09-702-705-1812
Sequence 1812, Application US/09702705
Patent No. 6504010
GENERAL INFORMATION:

APPLICANT: Wang, Tongtong
APPLICANT: Bangur, Chaifanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedvick, Tom
APPLICANT: Carter, Darlick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane

APPLICANT: Fan, Liqun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.478C14
CURRENT APPLICATION NUMBER: US/09/702.705
NUMBER OF SEQ ID NOS: 1833
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1812
LENGTH: 474
TYPE: PRT
ORGANISM: Homo sapiens
US-09-702-705-1812

Query Match 3.7%; Score 142.5; DB 4; Length 474;
Best Local Similarity 39.0%; Pred. No. 0.00097;
Matches 41; Conservative 10; Mismatches 25; Indels 29; Gaps 5;

```

QY 6 HGGGGGGSGPSAGSGGGGFGGSA-----AVALATASGKS-----GGGS 45
   ||||| : : : : : : : : : : : : : : : : : : : : :
Db 166 HGGGGGGSSNAGGGGAGSGGANSKPAOKKSGSVAGGAGGVSKPHAKLILAGG- 224
QY 46 CGGGGSYSASSSSSAAAAAG-AAVLVYKPKRMEHVADHLEFLQA 89
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 225 -GGKRAAAAAASFPAEDAGAAALLPL-----GAADHHSLYKA 262

```

RESULT 7

US-09-736-457-1812
Sequence 1812, Application US/09736457
Patent No. 6509448
GENERAL INFORMATION:

APPLICANT: Wang, Tongtong
APPLICANT: Bangur, Chaifanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedvick, Tom
APPLICANT: Carter, Darlick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
APPLICANT: Fan, Liqun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.478C15
CURRENT APPLICATION NUMBER: US/09/736.457
CURRENT FILING DATE: 2000-12-13
NUMBER OF SEQ ID NOS: 1864
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1812
LENGTH: 474
TYPE: PRT
ORGANISM: Homo sapiens
US-09-736-457-1812

Query Match 3.7%; Score 142.5; DB 4; Length 474;
Best Local Similarity 39.0%; Pred. No. 0.00097;
Matches 41; Conservative 10; Mismatches 25; Indels 29; Gaps 5;

```

QY 6 HGGGGGGSGPSAGSGGGGFGGSA-----AVALATASGKS-----GGGS 45
   ||||| : : : : : : : : : : : : : : : : : : : : :
Db 166 HGGGGGGSSNAGGGGAGSGGANSKPAOKKSGSVAGGAGGVSKPHAKLILAGG- 224
QY 46 CGGGGSYSASSSSSAAAAAG-AAVLVYKPKRMEHVADHLEFLQA 89
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 225 -GGKRAAAAAASFPAEDAGAAALLPL-----GAADHHSLYKA 262

```

RESULT 8

US-09-059-584-51
Sequence 51, Application US/09059584
Patent No. 6440701
GENERAL INFORMATION:

```

; APPLICANT: Myers, Lisa E
; APPLICANT: Schryvers, Anthony B
; APPLICANT: Hartness, Robin E
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Du, Run-Pan
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Kiehn, Michel H
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Slim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/059,584
; FILING DATE: 14-APR-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/778,570
; FILING DATE: 03-JAN-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24973
; REFERENCE/DOCKET NUMBER: 1038-794
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 712 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-09-059-584-51
;
; Query Match
; Best Local Similarity 19.8%; Score 142.5; DB 4; Length 712;
; Matches 100; Conservative 76; Mismatches 169; Indels 161; Gaps 20;
;
; Oy 9 GGGGGGSGP-----SAGSGGGGFGGSAVAATAAGSGSGGGGSGGGSYASSSS 58
;      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
;      22 GGGGSGNPAPPIPPNAGAGNAGSGTGN-ANAGSGGASGSGT--SAST-- 75
;
; Oy 59 SAAAGAALVPRKPKMEHQADHEL--FLAFKPTQIYRP-LRTNLIAPIFLHRL 115
;      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
;      76 -----QKPKYDVPPTDKKKKKEVSGIOEPAMGYELLRMTPO----- 115
;
; Oy 116 TMSHRNSTNIRKRTFYDMLSKYEKKKGEQESHLSAHLQITFTGFHKNDKPSPS 175
;      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
;      116 EDEHAKINTN-----DVYKLEGLKHN-----PFDSN 143
;
; Oy 176 --ENQNSATYLELVKVCCHKRKADYSCIRQVPTGKKOPLPLDINQTRPGNFPPLAYS 233
;      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
;      144 IMQNTKSKYQV---VYNQKQNTENQIKKENKELDKTALKALEKYLDDLTSLAMP 199
;
; Oy 234 SNEFSPNSHMVKSYSLEFRVTRPGRRENGMINGTENIDVNEELPARRRKNRREDGK 293
;      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
;      200 IYEKNTDNDHKONKAR---TRDLKYVSGIYRSGYSNIDOKTI----- 242
;
; Oy 294 TEVAQMTVEFDKNNRIQLDGEYEVAMQEMEBCPIS--KRAATWETLIDGRRLPPEFTFSQ 351
;      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
;      243 ----ANTGPD-----GALFYKGTQYAKQLPVSSEVYKGTWDFMTDAKKGGSFSFER 290
;
; Oy 352 GPTLQPTLWMTGE---TNDKS-----TAPIAKPLATRNSESLHDEKRP 391
```

```

;
; Db 291 RAGDRYSAMSHPEPSLITDDKKNPDNYDEXSHSEFTVDESKSLTGLFSNLQDHHK 350
;
; Oy 392 GSVKPTQITAV-----RESLTTLQTKKEDTRENQKXR 427
;      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
;      Db 351 KVTYKTRVDINARIHGNRPGSATAIKNQESKAKHPTSDADNLECGFGPNAEELA 410
;
; Oy 428 IFQFELYNNN-----TRQTEARD 446
;      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
;      Db 411 --GKFLTDNKLKLGVGAKQESAKE 434
;
;
; RESULT 9
; US-09-052-995-1
; Sequence 1, Application US/09052995
; Patent No. 6183956
; GENERAL INFORMATION:
; APPLICANT: SivaraJa, Mohanram
; APPLICANT: Strulovici, Berta
; APPLICANT: Flores, Osvaldo A.
; TITLE OF INVENTION: High Throughput In Vitro Screening Assay
; TITLE OF INVENTION: for Transcription Modulators
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/052,995
; FILING DATE: 31-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parent, Annette S.
; REGISTRATION NUMBER: 42,058
; REFERENCE/DOCKET NUMBER: 018781-000600US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 201 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
;
; US-09-052-995-1
;
; Query Match
; Best Local Similarity 3.6%; Score 141; DB 3; Length 201;
; Matches 26; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
;
; Oy 7 GGGGGGSGPSAGSGGGGFGGSAVAATAAGSGSGGGGSGGGSYASSSSAAAAAG 65
;      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
;      Db 92 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 150
;
;
; -RESULT 10
; US-09-053-003-40
; Sequence 40, Application US/09053003
; Patent No. 6207391
; GENERAL INFORMATION:
; APPLICANT: Wu, Pengguang
; APPLICANT: McKliney, Judi
; TITLE OF INVENTION: High-Throughput Screening Assays for
; TITLE OF INVENTION: Modulators of STAT4 and STAT6 Activity
```

NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/053,003
FILING DATE: 31-MAR-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Smith, Timothy L.
REGISTRATION NUMBER: 35,367
REFERENCE/DOCKET NUMBER: 018781-000800US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 201 amino acids
TYPE: amino acid
STRANDEDNESS:
MOLECULE TYPE: linear
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1..97
OTHER INFORMATION: /product- "OTHER"
OTHER INFORMATION: /note- "Gly at positions 1-97 may be
OTHER INFORMATION: present or absent"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 105..201
OTHER INFORMATION: /product- "OTHER"
OTHER INFORMATION: /note- "Gly at positions 105-201 may be
OTHER INFORMATION: present or absent"
US-09-053-003-40
Query Match 3.6%; Score 141; DB 3; Length 201;
Best Local Similarity 44.1%; Pred. No. 0.00035;
Matches 26; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
Oy 7 GGGGGGSPAGSGGFGGSAVAAATASGKSGGCGGGSYSNSSSSAAAAAG 65
Db 92 GCGGGGGGPGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 150
RESULT 11
US-09-054-281-22
Sequence 22, Application US/09054281
Patent No. 6444421
GENERAL INFORMATION:
APPLICANT: Chung, Jay H.
TITLE OF INVENTION: Methods for Detecting Intermolecular
TITLE OF INVENTION: Interactions In Vivo and In Vitro
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/054,281
FILING DATE: 02-APR-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/080,234
FILING DATE: 03-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/826,622
FILING DATE: 03-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Smith, Timothy L.
REGISTRATION NUMBER: 35,367
REFERENCE/DOCKET NUMBER: 015280-295200US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 201 amino acids
TYPE: amino acid
STRANDEDNESS:
MOLECULE TYPE: linear
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1..97
OTHER INFORMATION: /product- "OTHER"
OTHER INFORMATION: /note- "Gly at positions 1-97 may be
OTHER INFORMATION: present or absent"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 105..201
OTHER INFORMATION: /product- "OTHER"
OTHER INFORMATION: /note- "Gly at positions 105-201 may be
OTHER INFORMATION: present or absent"
US-09-054-281-22
Query Match 3.6%; Score 141; DB 4; Length 201;
Best Local Similarity 44.1%; Pred. No. 0.00035;
Matches 26; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
Oy 7 GGGGGGSPAGSGGFGGSAVAAATASGKSGGCGGGSYSNSSSSAAAAAG 65
Db 92 GCGGGGGGPGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 150
RESULT 12
US-09-478-948-6
Sequence 6, Application US/09478948
Patent No. 6465258
GENERAL INFORMATION:
APPLICANT: Shan, Bei
APPLICANT: Okamoto, Arthur Y.
APPLICANT: Tularik Inc.
TITLE OF INVENTION: FXR Receptor-Mediated Modulation of Cholesterol
TITLE OF INVENTION: Metabolism
FILE REFERENCE: 018781-001310US
CURRENT APPLICATION NUMBER: US/09/478,948
CURRENT FILING DATE: 2000-01-06
PRIOR APPLICATION NUMBER: US 60/115,249
PRIOR FILING DATE: 1999-01-07
NUMBER OF SEQ ID NOS: 6
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 6
LENGTH: 201
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: flexible linker
NAME/KEY: MOD_RES

LOCATION: (1)..(97)
OTHER INFORMATION: gly at positions 1-97 may be present or absent
NAME/KEY: MOD.RES
LOCATION: (105)..(201)
OTHER INFORMATION: gly at positions 105-201 may be present or absent
US-09-478-948-6

Query Match 3.6%; Score 141; DB 4; Length 201;
Best Local Similarity 44.1%; Pred. No. 0.00035;
Matches 26; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 7 GGGGGGGSPSGSGGGGGGSAVAATAAGKSGSGGGGSGTSSSSSSAAAAAG 65
DB 92 GGG 150

RESULT 13
US-07-789-915A-8
Sequence 8, Application US/07789915A
Patent No. 5212058

GENERAL INFORMATION:
APPLICANT: Baker, Rohan T.
APPLICANT: Tobias, John W.
APPLICANT: Varshavsky, Alexander
TITLE OF INVENTION: Ubiquitin-Specific Proteases
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02173

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/789,915A
FILING DATE: 19911108
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E.
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: MIT-5091AA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540

INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 912 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-789-915A-8

Query Match 3.6%; Score 139.5; DB 1; Length 912;
Best Local Similarity 18.1%; Pred. No. 0.0048;
Matches 141; Conservative 107; Mismatches 264; Indels 269; Gaps 31;

QY 28 SAAVAATAAGKSGSGGGGSGGSGYSSSSSSAAAAAVALPKPKMEHVQADHELFL 87
DB 102 SNGITNNNGSSGNGGSSGSGSYKSKHT----- 131
QY 88 QAEPKPTQIYRFLRPNLAPFLAHTLTYSMSRNSRTIKRRTFVVDMLSFYKMGKE 147
DB 132 -----YHHYNSNNHI-----PMASPNNGSN-----AGMKKOTNS 161
QY 148 QESHSAHLQFLTFEFGKNDKSPNSEENQSVTLVLLVYVCKHKKRDVSCPIROY 207
DB 162 SNGGSSA-----TSPSYSTYSSSSOYD-----LYKFDVTKLN-----LKENS 200

QY 208 TGRKOVPLIPDLNQRPG-----NPSLAVSSNEFEPSPN-----SHWKSYSGL 250
DB 201 SMLIDLPLE--INTTEAEFAAASVORYELNMMALNINSELSSEKSSAAHHTKSHSI 258
QY 251 LFRVTRPGRRFENGMIINGETN--ENIDVNEELPARRKRNDEDEKTFVQMTYFDKRRQLQ 309
DB 259 -----PKHNEEYKTTGHEEDAHKDPHASKDAHELKKEVKK----- 298
QY 310 LLDGEYVAMQMEECPISKKRATWETLIDGRLPFEFFSOGPTLQFLRTGTETNDS 369
DB 299 -----EDAKODRNEVIOEPQATVLPYVDKKE--PEEVEEN-----TSKTSSSS 341
QY 370 TAPIA---KPLATRNSESLHGENKPGSVKPTQTIAVKSLLTDTQTRREKQTPMENRQK 425
DB 342 PSPPAKSWSAIADAIKSRQASNKTVSGSMVTKPISGTFAGVSTNMAAATIGRSSSP 401
QY 426 LRIFFQFLYNNNTROOTEARDDLHCWCTLNCRLKLSLKLKCHSRITFNVTYHPRKA 485
DB 402 L-----LSKPPQKDKKYYPPSTKGIEPLGSLA--LRMCDDPPISTVILNKKV 448
QY 486 RIDVSINECYDGSYAGNPQDIHRQPGFAFSRNGPVKRTPIYHLVCRP-----KR 535
DB 449 ENKIPVHSII-----PGIIRANICF-----MSVQLVLLYCKRPTDIVYINVLTNRN 495
QY 536 TRASM-----SEPLESDG--EVEQORTYSSGHNRLYFHS 569
DB 496 TNSRVGTSSCKLIDACTLTYKQFDKETEYKFLFNDDAKETTESDAKSSKSKSPQCHA 555
QY 570 TGLPLRQEMENDSEDEKPEMLREKTIQIEEFSVNGEKEVMTLMLHYKHK----- 624
DB 556 TADAVKPDFE-----YKTLSTIPKFKDLQWGHQDEAFETHLLDQJHEEL 601
QY 625 -----GFIADONMHACMLFVENYQKRIKKNLCRNFMHLVSMH---DENLISIMSIDRA 677
DB 602 ISALDGLTNE-----IQNMLQSIINDEQL-KYFIFNLISYKGAEP-----IKNA 645
QY 678 VTKLRMQOKLEKESASPANEETEEQONTANGFSEINSKERALETDSVSGYSKQSKRO 737
DB 646 SPRLKELIKRY-----GVINDSTEE-----NGMHE-----VSGSKRGRKT 682
QY 738 K 738
DB 683 K 683

RESULT 14
US-08-005-002C-8
Sequence 8, Application US/08005002C
Patent No. 5494818

GENERAL INFORMATION:
APPLICANT: Baker, Rohan T.
APPLICANT: Tobias, John W.
APPLICANT: Varshavsky, Alexander
TITLE OF INVENTION: Ubiquitin-Specific Proteases
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Kevin M. Farrell, P.C.
STREET: P.O. Box 999
CITY: York Harbor
STATE: Maine
COUNTRY: U.S.A.
ZIP: 03911

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/005,002C
FILING DATE: 15-JAN-1993
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER: US 07/789,915

```

FILING DATE: 08-NOV-1991
ATTORNEY/AGENT INFORMATION:
NAME: Farrell, Kevin M.
REGISTRATION NUMBER: 35,505
REFERENCE/DOCKET NUMBER: MIT-5091AAZ
TELECOMMUNICATION INFORMATION:
TELEPHONE: 207-363-0558
TELEFAX: 207-363-0528
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 912 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-005-002c-8

Query Match      3.68; Score 139.5; DB 1; Length 912;
Best Local Similarity 18.18; Pred. No. 0.0048;
Matches 141; Conservative 107; Mismatches 264; Indels 269; Gaps 31;

QY 28 SAAVAATAAGSGKGGCGGSGSYSASSSSAAAAAGAAVLPVKPKMEHVQADHLEFL 87
D 102 SNGITNNNGSSGNGANGSSGSMYKNSHT----- 131
QY 88 QAFKPTQIYRFLTRNLAPIFLHRTLTYSHRNSRTNKRKTFKYVDMLSVEKKKE 147
D 132 -----YHNHNSNNHI-----PMWASPNSSGN-----AGKKQOTNS 161
QY 148 QESHLSAHLQLTFTGFHKNDKPSNSEONSVTLVLLVKYCHKKRKDVSCPIQVP 207
D 162 SNGGSSA-----TSPSYSSYSSSOYD-----LYKDVYTKLN-----LKENS 200
QY 208 TGKQVLPIDLNQTKG-----NPSLAVSNEPEPSN-----SHWVKSYSL 250
D 201 SNIQLPLF--INTTEAFPAASVQRYELNKKALNLSSELSSESVKSSAHHTKSHSI 258
QY 251 LFRVTRGRREFNGINGETN-ENIDVNEELPARKKRNRREDGEKTFVAQMTVPDKNRLQ 309
D 259 -----PKHNEEVKTEHGEEDAHDKKPHASKDAHLEKKTEYVK----- 298
QY 310 LLDGEYVAQMEECPISKKRATWETILDGKRLPPEPESQGPFTLOFTLRMTGETNDS 369
D 299 -----EDAKQDRNEKVIQEPQATVLPYVDKE--PESVSEEN-----TKTSPPS 341
QY 370 TAPLA-----KPLATRNESLHQENKPGSVKFTQTIAVKESITLTDQTRKEKDPENRQK 425
D 342 PSPPAAKSWSAIASDAKRSQASNKTVSGSWVTKTPISGTYAGVSSTNMAAATIGKSSSP 401
QY 426 LRITQFLYNNNTQOETARDLCHPWCCTLNCRKLYSLKLKXCHSRFIFNYYYHKG 485
D 402 L-----LSKQPKKDKKTYPPSTKGIPLGSLA--LRKCFDPDFTSYLRMKDV 448
QY 486 RIDVINECYDGYAGNPDIHROPGFASRNGFVKRTPIYHILVCRP-----KR 535
D 449 ENKIPVHSII-----PGIIRNANICF-----MSVLYQVLLCYCFPIDIVINVLSTRN 495
QY 536 TKAM-----SFELESDG--EVEQQRYSSSHNLNLYRSD 569
D 496 TNSVAVGSSCKLLDACLTMYKQFDKETEYKFKELNADAEKTTSDAKSKSSKSFQCA 555
QY 570 TCLPLRFOEMEVDEDEKDEPEMLREKTIITOIEEFDVNEGEKVKMLNLYVMH----- 624
D 556 TADAVKDEF-----YKTLSTIPKFKDLOMGHQAEEFLHLLQQLHEEL 601
QY 625 -----GTADQNMNHACHLFEVNTYQKIIKKNLCSNMLHLVSMH--DPLNLSIMSIDKA 677
D 602 ISAIIDLTVNE-----IQNMLOSINDEQL-KVFIFRNLNRYGKAEE-----IKNA 645
QY 678 VTKIREMOQLEKESASPANEETTEBONGTANGFSEINSEKALETDSVSGVSKQSKQ 737
D 646 SPRLKELIEKY-----GVINDSTEE--NGMHE-----VSGSSKRGKKT 682
QY 738 K 738

```

```

Db 683 K 683

RESULT 15
US-08-487-203A-8
Sequence 8, Application US/08487203A
Patent No. 5683904
GENERAL INFORMATION:
APPLICANT: Baker, Rohan T.
APPLICANT: Tobias, John W.
APPLICANT: Varshavsky, Alexander
TITLE OF INVENTION: Ubiquitin-Specific Proteases
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kevin M. Farrell, P.C.
STREET: P.O. Box 999
CITY: York Harbor
STATE: Maine
COUNTRY: U.S.A.
ZIP: 03911
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,203A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/005,002
FILING DATE: 15-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Farrell, Kevin M.
REGISTRATION NUMBER: 35,505
REFERENCE/DOCKET NUMBER: MIT-5091A32
TELECOMMUNICATION INFORMATION:
TELEPHONE: 207-363-0528
TELEFAX: 207-363-0528
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 912 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-487-203A-8

Query Match      3.68; Score 139.5; DB 1; Length 912;
Best Local Similarity 18.18; Pred. No. 0.0048;
Matches 141; Conservative 107; Mismatches 264; Indels 269; Gaps 31;

QY 28 SAAVAATAAGSGKGGCGGSGSYSASSSSAAAAAGAAVLPVKPKMEHVQADHLEFL 87
D 102 SNGITNNNGSSGNGANGSSGSMYKNSHT----- 131
QY 88 QAFKPTQIYRFLTRNLAPIFLHRTLTYSHRNSRTNKRKTFKYVDMLSVEKKKE 147
D 132 -----YHNHNSNNHI-----PMWASPNSSGN-----AGKKQOTNS 161
QY 148 QESHLSAHLQLTFTGFHKNDKPSNSEONSVTLVLLVKYCHKKRKDVSCPIQVP 207
D 162 SNGGSSA-----TSPSYSSYSSSOYD-----LYKDVYTKLN-----LKENS 200
QY 208 TGKQVLPIDLNQTKG-----NPSLAVSNEPEPSN-----SHWVKSYSL 250
D 201 SNIQLPLF--INTTEAFPAASVQRYELNKKALNLSSELSSESVKSSAHHTKSHSI 258
QY 251 LFRVTRGRREFNGINGETN-ENIDVNEELPARKKRNRREDGEKTFVAQMTVPDKNRLQ 309
D 259 -----PKHNEEVKTEHGEEDAHDKKPHASKDAHLEKKTEYVK----- 298
QY 310 LLDGEYVAQMEECPISKKRATWETILDGKRLPPEPESQGPFTLOFTLRMTGETNDS 369

```

```
Db 299 -----EDAKODRNEKVIQEPQATVLPVYDKRE--PEESVEEN-----TSKTSPPS 341
QY 370 TAPIA-----KPLATRSESLHOENKPGSVKPTQIYAVKESL/TDLOTRKENDPNNRQK 425
Db 342 PBPRAKMSAIAISDAIKSRQASNKTVSGSMWTKTPTISGTTAGVSSTNMAAATIGKSSP 401
QY 426 LRIYQFLYNNNTROOTEARDDLHCPCWCTLNCRLYSLKHLKLSHREIFNYVYHPKGA 485
Db 402 L-----LSKQPKKKKKYVPPSRKGIEPLGSLA--LMCFDPDFISYVLANKDY 448
QY 486 RIDVSIINECYDGSYAGNPQDIIHRQGFAPSRNGPVKRPITHTLVCRP-----KR 535
Db 449 ENKIPVHSII-----PRGIINRANICF-----MSVLYQVLLYCKPFDIVYNLSTRN 495
QY 536 TKASM-----SEPLESDG--EVRQORTYSSGHNRLYFHS 569
Db 496 TNSRVGTSSCKLIDACLIMYKQFDKETYERKFLFNADDAKKTESDAKSSKSKSFQHC 555
QY 570 TCLPLRPOEMEYDSEDEKDPWMLREKTIQIEEFDVNEGEKEVMKLMNLHVAKH----- 624
Db 556 TADAVRPDEF-----YKTLSTIPKFKDLQNGCHQDAEFLTHLDQLHEL 601
QY 625 ----GFADNQNHNHACMLFVENTGQKIIKKNLGRNFMHLVSMH---DENLISIMSIDRA 677
Db 602 ISALDGLTNE-----IONMLQSIINDEQL-KVFIRNLSRYGKAEF-----IKNA 645
QY 678 VTKLRMOOKLEKESASAPANEITEEQNGTANGFSEINSKEKALETDSVSGVSKOSKQ 737
Db 646 SPRLKELIERKY-----GYINDSTEE---NGWHE-----VSGSSKRGKKT 682
QY 738 K 738
Db 683 K 683
```

Search completed: September 2, 2003, 11:06:56
Job time : 26.3289 secs

THIS PAGE BLANK (USPTO)


```

Db 125 AAAAGAALVPKPKMEHVOADHELFLQAEKPTQIYRFLRNLIAPIFLHRTLYMSH 184
QY 121 RNSRTIKRRTFEVDMLSKVEKMGEOESHSLSAHLQLETFGFFHKNDKPSPSNEON 180
    185 RNSRTIKRRTFEVDMLSKVEKMGEOESHSLSAHLQLETFGFFHKNDKPSPSNEON 244
QY 181 SYTLEVLAVKCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTPGNFPLAASNEEPS 240
    245 SYTLEVLAVKCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTPGNFPLAASNEEPS 304
QY 241 NSHMYVSYSLFVTPRGREFNGMNGETNENIDVNEELPARKKNNREDGEKTFVAOMT 300
    305 NSHMYVSYSLFVTPRGREFNGMNGETNENIDVNEELPARKKNNREDGEKTFVAOMT 364
QY 301 VFDKNRRLQDLDEYEVAQOMECEPISSKRATWETILDGKRLPPETFSOGPTLOFTLR 360
    365 VFDKNRRLQDLDEYEVAQOMECEPISSKRATWETILDGKRLPPETFSOGPTLOFTLR 424
QY 361 WTGETNDKSTAPIAKLATRNSESLHQENKPGSVKPTQITIAVESLTTDLQTRKEDTPN 420
    425 WTGETNDKSTAPIAKLATRNSESLHQENKPGSVKPTQITIAVESLTTDLQTRKEDTPN 484
QY 421 ENROKLRIFYQFLYNNNTROOTEARDLHCPCMTLNCRLKLSLKLKLCCHSFIFINYY 480
    485 ENROKLRIFYQFLYNNNTROOTEARDLHCPCMTLNCRLKLSLKLKLCCHSFIFINYY 544
QY 481 HPRGARIDVSYNCTGYSAGNPQDIHRQGFASRNGPKRPTIHLILVCRKPKTKASM 540
    545 HPRGARIDVSYNCTGYSAGNPQDIHRQGFASRNGPKRPTIHLILVCRKPKTKASM 604
QY 541 SEFLSEDEGEVEQORTYSSGHNRLYFHSPTCLPLRQEMEVSDEKDEPMLEKTIITQI 600
    605 SEFLSEDEGEVEQORTYSSGHNRLYFHSPTCLPLRQEMEVSDEKDEPMLEKTIITQI 664
Db 601 EEFSDVNEGEKEMKLMNLHYMKHGFADQNMNHACMLEVENYGOKIIRKNCRNMLHL 660
    665 EEFSDVNEGEKEMKLMNLHYMKHGFADQNMNHACMLEVENYGOKIIRKNCRNMLHL 724
QY 661 VSMHDFNLISIMSIDKAVTKLRMOOKLEKESASPANEITBEONGTANGFSEINSKER 720
    725 VSMHDFNLISIMSIDKAVTKLRMOOKLEKESASPANEITBEONGTANGFSEINSKER 784
Db 721 ALETDSVSGVSKOSKROKL 739
    785 ALETDSVSGVSKOSKROKL 803

```

RESULT 2

096BD9 PRELIMINARY: PRT: 739 AA.

AC 096BD9; 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE joined to Jazf1.

OS Homo sapiens (human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RA [1]

RP SEQUENCE FROM N.A.

RA STRAUSBERG R.;

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC015704; AAH15704.1; -

DR InterPro; IPR007087; Znf_C2H2.

DR SMART; SM00355; Znf_C2H2.1.

DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.

SO SEQUENCE 739 AA; 83054 MW; A8830EBC3FD38D56 CRC64;

Query Match 99.84; Score 3878; DB 4; Length 739;

Best Local Similarity 99.94; Pired. No. 3.9e-263;

Matches 738; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 MAPQKHGGGGGGGSGSAGSGGGGFGGSAVAATAAGSGKSGGGGSGGGSYASASSSA 60
    1 MAPQKHGGGGGGGSGSAGSGGGGFGGSAVAATAAGSGKSGGGGSGGGSYASASSSA 60
Db 61 AAAAGAALVPKPKMEHVOADHELFLQAEKPTQIYRFLRNLIAPIFLHRTLYMSH 120
    61 AAAAGAALVPKPKMEHVOADHELFLQAEKPTQIYRFLRNLIAPIFLHRTLYMSH 120
QY 61 SYTLEVLAVKCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTPGNFPLAASNEEPS 240
    61 SYTLEVLAVKCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTPGNFPLAASNEEPS 240
QY 121 RNSRTIKRRTFEVDMLSKVEKMGEOESHSLSAHLQLETFGFFHKNDKPSPSNEON 180
    121 RNSRTIKRRTFEVDMLSKVEKMGEOESHSLSAHLQLETFGFFHKNDKPSPSNEON 180
QY 181 SYTLEVLAVKCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTPGNFPLAASNEEPS 240
    181 SYTLEVLAVKCHKKRKDVSCPIRQVPTGKKQVPLIPDLNQTPGNFPLAASNEEPS 240
QY 241 NSHMYVSYSLFVTPRGREFNGMNGETNENIDVNEELPARKKNNREDGEKTFVAOMT 300
    241 NSHMYVSYSLFVTPRGREFNGMNGETNENIDVNEELPARKKNNREDGEKTFVAOMT 300
QY 301 VFDKNRRLQDLDEYEVAQOMECEPISSKRATWETILDGKRLPPETFSOGPTLOFTLR 360
    301 VFDKNRRLQDLDEYEVAQOMECEPISSKRATWETILDGKRLPPETFSOGPTLOFTLR 360
QY 361 WTGETNDKSTAPIAKLATRNSESLHQENKPGSVKPTQITIAVESLTTDLQTRKEDTPN 420
    361 WTGETNDKSTAPIAKLATRNSESLHQENKPGSVKPTQITIAVESLTTDLQTRKEDTPN 420
QY 421 ENROKLRIFYQFLYNNNTROOTEARDLHCPCMTLNCRLKLSLKLKLCCHSFIFINYY 480
    421 ENROKLRIFYQFLYNNNTROOTEARDLHCPCMTLNCRLKLSLKLKLCCHSFIFINYY 480
QY 481 HPRGARIDVSYNCTGYSAGNPQDIHRQGFASRNGPKRPTIHLILVCRKPKTKASM 540
    481 HPRGARIDVSYNCTGYSAGNPQDIHRQGFASRNGPKRPTIHLILVCRKPKTKASM 540
QY 541 SEFLSEDEGEVEQORTYSSGHNRLYFHSPTCLPLRQEMEVSDEKDEPMLEKTIITQI 600
    541 SEFLSEDEGEVEQORTYSSGHNRLYFHSPTCLPLRQEMEVSDEKDEPMLEKTIITQI 600
Db 601 EEFSDVNEGEKEMKLMNLHYMKHGFADQNMNHACMLEVENYGOKIIRKNCRNMLHL 660
    601 EEFSDVNEGEKEMKLMNLHYMKHGFADQNMNHACMLEVENYGOKIIRKNCRNMLHL 660
QY 661 VSMHDFNLISIMSIDKAVTKLRMOOKLEKESASPANEITBEONGTANGFSEINSKER 720
    661 VSMHDFNLISIMSIDKAVTKLRMOOKLEKESASPANEITBEONGTANGFSEINSKER 720
Db 721 ALETDSVSGVSKOSKROKL 739
    721 ALETDSVSGVSKOSKROKL 739

```

RESULT 3

09VW55 PRELIMINARY: PRT: 855 AA.

AC 09VW55; 08T9D8; 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE CG8013 protein (SD04959P).

GN SU(2)12 OR CG8013.

OS Drosophila melanogaster (fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RA [1]

RP SEQUENCE FROM N.A.

RA STRAIN=Berkeley;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Asdnbner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H.C., Blake R.G., Chame M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Barker E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Bailey R.M., Basu A., Baxendale J., Bayraktaroglu U., Besley E.M.,
 RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
 RA Borotova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,
 RA Durkin K.D., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glaeser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jabali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Ruskern D.R., Pacled J.M.,
 RA Palazolo M., Pitman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Slden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spler E., Spreading A.C., Stapleton M., Strong R., Sun E.,
 RA Sylskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT *The genome sequence of Drosophila melanogaster.*;
 RL Science 287:2185-2195(2000).
 RP
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Barzon J., An H., Baldwin D., Barzon J., Beeson K.Y., Busan D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreira S., Frise E., Galle R.F., Gary N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jabali M., Kruse D., Li P., Maitel B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacled J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pitman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Sylskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT *Sequencing of Drosophila melanogaster genome.*;
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN
 RN SEQUENCE FROM N.A.
 RP
 RP Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN
 RN SEQUENCE FROM N.A.
 RP
 RP Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.

RN [6]
 RP SEQUENCE FROM N.A.
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Chame M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guatin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nunoo J., Pacled J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
 RL Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AE003515; AAF4904.2;
 DR EMBL: AY069809; AAL39954.1;
 DR Flybase: FBgn020887; Slc212.
 DR InterPro: IPR007087; Znf_C2H2.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1;
 SQ SEQUENCE 855 AA; 95316 MW; C10FFC6013954105 CRC64;
 Query Match 30.3%; Score 1177; DB 5; Length 855;
 Best Local Similarity 34.8%; Pred. No. 6, 8e-74;
 Matches 281; Conservative 136; Mismatches 225; Indels 166; Gaps 26;
 43 GSGCGG--GGSYASSSSSSAAAGAAVLPK--RPKMEHQADHFLQAFEPPTQIR 98
 15 GSAANGIIGLTGAPDASN---AGSTVPTAEQGVKLGHQEGDELFLQAFEPPTQIR 70
 99 FRTNRLNLPITLRLITLYMSRNRFTNRKRTFYVDMLSKVEKKGEQESHLSA-HL 157
 71 YLRNRETPIDPLNLTSLYKERMSSNNKRISFQVNSLSEI-----TQSEAVSONYL 125
 158 QLTGTFEFKNDKPPNSNEQ-----NSTLEVLVYCKRKRKQVSPICQV 206
 126 HVIYSLHKKLPARLDNSEGEDLLOQLCEGSESVETLYTKTRSRKSTQDFOFL 185
 207 PYGKROVPLIPDLNQTGPNPPLAVSSNEEP--SNHVKYSYSLFRTGRDREPNM 265
 186 LKCSQIYVNP---KDRVEHNTISIPLOTMRPMEOHLL--YKLLFRK-----V 231
 266 INGFENENIDVBEELPARKRRNDEGEKTFVAQMTVEKRRRLQDLGSEYVAMQMEEC 325
 232 LSPSTCN--DENAETPP--NKRSPN--EKMFGESELLYKRSQF--ITEGEYEMLPINST 286
 326 PI---SKKRAWETLIDGRLP---PEFEFSQPTLOFLMTGTENDSTAPIKPLAT 379
 287 SIKSSPKKCTWEIWD--SYPLSLTYDYQOSPMLKFLTL--SNEOLPMETAPELQ 342
 380 R-----NSESILQEN-----KPSGVKPTQTAVRESLTTDQTRKREKDPNE 421
 343 RYVQHLDAVAEMNNNNNNNNNNNSGLKNGSGGGSYTC-----KTPP-- 386
 422 NRQKLRIFQPLNNNTROQTARDDLCPWCTLNCRLKYSILKHLKLCSPFTENYVH 481
 387 --EHQIYVNFYSNNTRQOTETQELNCPMGCLCLRLYALKHLKLCSPFTENYVH 444
 482 PKGARIVSINCYGSGVGNPNODIHRQGFAPSN--GVKRTPTITLIVYCPKPKSM 540
 445 GSGARIDVTINATPOSTAGSPYDLGSGSFARCTGVKRTPTITLIVYCPKPKSM 504
 541 SEFLSEDEGEVQORTYSSGNRLYFHSPTCLPLRPOEMEVSEDEKDEPMTRETIQI 600
 505 DEFLDEDEDEISNGSYIGHNRVHTHTCTCPYRKELDIDSEBSPDMLRQTIQI 564
 601 EEPFSDVNEGEKVMKMLNLYKKGFIADONNHACMLFVENYGOKIIRKNCIRNPMHL 660
 565 DEFSVNEGEKELMKMLNLYKKGFIADONNHACMLFVENYGOKIIRKNCIRNPMHL 624
 661 VMHPDNLSINSIDKAVTKLREMOQKLEKGS----- 693
 625 GSLFDGLTAELVTKVTKVTKVTKVTKVTKVTKVTKVTKVTKVTKVTKVTKVTKVTK 684
 694 -----ASPANEETITE-----EONGTANG 711
 685 LKSPQKPPADQASSTASSTSGSGSSSMQPPKRPMAHLKRGSAASSPGVQSGTENG 744
 712 FSEINSKEKALETDSYGVSKOSKROKL 739

Db 745 TNGSNS-----SSSNKNAKKSADPDL 767

RESULT 4

09NUG9 PRELIMINARY: PRT: 900 AA.

AC 09NUG9: 01-OCT-2000 (Tremblrel. 15, Created)

DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)

DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)

DE Su(2)12.

GN Su(2)12 OR CG8013.

OS Drosophila melanogaster (fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RA Blive A., Rasmuson-Leestander A., Larson J.;

RT *Suppressor of zeste 12.*

RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF149047; AAF73149.1; -

DR FlyBase: FBgn0020887; Su(2)12.

DR InterPro: IPR007087; ZNF_C2H2.

DR SMART: SM00355; ZNF_C2H2; 1.

DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 1.

SQ SEQUENCE 900 AA; 100103 MW; 53BA0D83C49EC92F CRC64;

Query Match 30.38; Score 1177; DB 5; Length 900;

Best Local Similarity 34.88; Pred. NO. 7.2e-74;

Matches 281; Conservative 136; Mismatches 225; Indels 166; Gaps 26;

QY 43 GSGCG--GGSYSASSSSAAAGAAVLVVK--KPMHVOADHLELQAFKPTQYR 98

15 GSANGITGLHGAPOASN---AGSTVPTAGOVKLNHQOQELFLQAFKPTQYR 70

QY 99 FLKRNIAPIFLHRTITLYSHNSRNIRKTKFYVDMLSKYEKKGEDESHSLA-HL 157

71 YLRNRHETNPILFLKRLTSYKMERMRNKRKRISFOVSMESI-----TKSEAVSONYL 125

QY 158 OLPTGTFHNDKRPSENDQ-----NSVLELVLVVCHKRRDVSCTPIROY 206

126 HVIYDSJHEKLPALNDESGEDLOQLCEAGESVSEVTLTKRSKRKDSLDFOEL 185

QY 207 PTGKQVPLIPDLNQTGPNFPLAVSNEFER-SNSHWKSYSLFRVTRPGREFNGM 265

186 LSKSQQLVNP---KQVGEHATISIPLOIMRPGEOHTL--YKLLFRIR-----V 231

QY 266 INGTENNIDVNEELPARRKRNREDGKTVVAQMTVFDKNRRLQLDGEYEVANQMEEC 325

232 LSPSTCN--DENMETPP--NKRSRPN--EKMFGSELILEKSSGF--ITGEYEAMLIQPINST 286

QY 326 PI---SKKRAWTETILGKRLP---PRETFSGOGLTOFTLRMTGETNDKSTAPRLAT 379

287 SIFSFSKCKTWTETMPD-STIPLSLTYDYVQOSFMLEFHLTL---SNEQLPEMISAELO 342

QY 380 R-----NSBSLHQEN-----KPGSVKPTOTIAVKSLLTDTLRKEKDPNE 421

343 RYVCHLDVAVEMNNNNNNNNNNCSGLKNGSGGSGNTVC-----KTRP-- 386

QY 422 NRQKLRTFYQFLYNNNRTOOTEARDDHCPWCTLNCKKLYSLKHLKLSRFFENYVH 481

387 --EHIQIVNFMYSNNRTOOTEYQOELNCWCGIDCLRLALAKHLKCHARFFTYQPA 444

QY 482 PKGARIVSINECDYAGNPDIHQPGFAFSRN--GPVKRPTITILVCRPRTKASM 540

445 GSGARIVYTINDATDGSYASPYDLAAGPSSSFAKTCGPVAKTSVSLACRPRROTCCL 504

QY 541 SEPLESDGEVQOQRTYSSGHNRLYFHSIDCLPLRQEMEVNDSDEKDPMLREKTTQI 600

505 DEFLEDEDEISNQRSYITGHNRLYHNTETCLPVHPRFELIDSDGESDPLMLNQKTIOMI 564

QY 601 EEFSDVNEGEKYEYKMLNLMHVKHGFADNQNHNHACMLFEVNTGQIKKNCRNFMHL 660

565 DEFSVDNEGEKEKMLNLMHVMRHFVGDQCLPIACEMFLDAKGTETVFRNLYRNFIHM 624

QY 661 VSMHDFNLISIMSIDAVYTLREMOOKLEKES-----DQNTANG 711

625 CSLFDGLIAETVYTKVQKLGSLSKYAAGELMORQREOLKYWLDGMHKKQEDPKT 684

QY 694 -----ASPANEITE-----DQNTANG 711

685 LKSPQKRAPPADQASNTSSASTSGSGSSSQMPKPMHLLKRGSAASSPGVSKGTENG 744

Db 745 TNGSNS-----SSSNKNAKKSADPDL 767

QY 712 FSEINSEKALETDSVSGVSKQSKOKRL 739

RESULT 5

099L07 PRELIMINARY: PRT: 114 AA.

AC 099L07: 01-JUN-2001 (Tremblrel. 17, Created)

DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)

DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)

DE Hypothetical 12.8 kDa protein (fragment).

GN D11ERTD530E.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sclurognath; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA Strausberg R.;

RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: BC003922; AAH03922.1; -

DR MGD; MGI:1261758; D11ertd530e.

KW Hypothetical protein.

FT NON_TER

SQ SEQUENCE 114 AA; 12849 MW; A73CE641A1E9228 CRC64;

Query Match 14.28; Score 550; DB 11; Length 114;

Best Local Similarity 94.78; Pred. NO. 3.9e-31;

Matches 108; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 626 FIADNMNHAACMLFEVNTGQIKKNCRNFMHLVSMHDFNLISIMSIDAVYTLREMO 685

1 FIADNMNHAACMLFEVNTGQIKKNCRNFMHLVSMHDFNLISIMSIDAVYTLREMO 60

QY 686 OKLEKGSASPADEITEBONGTANGSEFNSKEKALETDSVSGVSKQSKOKRL 739

61 OKLEKGSASPADEITEBONGTANGSEFNSKEKALETDSVSGVSKQSKOKRL 114

Db 61 OKLEKGSASPADEITEBONGTANGSEFNSKEKALETDSVSGVSKQSKOKRL 114

RESULT 6

08L6Y4 PRELIMINARY: PRT: 626 AA.

AC 08L6Y4: 01-OCT-2002 (Tremblrel. 22, Created)

DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)

DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)

DE Hypothetical protein.

GN AT5G51230; AT5G51240.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Tripp M., Southwick A., Karlin-Neumann G., Nguyen M., Miranda M.,

RA Palm C.J., Bowser L., Jones T., Banh J., Carlinici P., Chen H.,

RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kanlaya A., Kawai J.,

RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,

RESULT 8

08WSB2 PRELIMINARY; PRT; 445 AA.
 AC 08WSB2;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DE Vernalization 2 protein.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosida II; Brassicales; Brassicaceae; Arabidopsids.
 NCBI_TaxID=3702;

Query Match 5.7%; Score 221; DB 10; Length 445;
 Best Local Similarity 21.1%; Pred. No. 2.6e-07;
 Matches 84; Conservative 64; Mismatches 151; Indels 100; Gaps 12;
 QY 413 RKEQDTNENRQKIRITQYQFLYNNNTQQTREARDLHCPCWCTLNCRLYSILKLKLSHS 472
 DB 58 RKRK-----SRSTGVVFNKDCNNTLQKTEVREDSCPCSMCGSFKGLQFHLNSSH 112
 QY 473 RFTFNYYHFKGARIDVSI-----NECYDG-----SYAGNPQDIIHROPGFARNG 518
 DB 113 LFEFEFLSEYQVNVSVKLSNFIIEEGSDDDKFEFSLCSKPRK-RKQRC---GRNN 168
 QY 519 PVKRTPTTHILVCPKRTKASMEFELESDG----- 549
 DB 169 -TRLKVCFLPLDSPSLTNGTENGITLLNDNGRGLGYPEATELAGQFEMTSNIPPAIHS 227
 QY 550 -----EVEQRTYSSGH-----NRLYFSDCLPLRPDEM--EV 581
 DB 228 SLDAKAVILTSEAVPATRTKLSAERSASRSHLLQKQFYSHRVPALAEQVMSDR 287
 QY 582 DSEDEKPEWLREKTIQIEEFSQVNGEKEVKMLNLMHYMKHGFADNOMNHACMLFEVE 641
 DB 288 DSEDEVDVADFEEDRMDDFDVYNDKDEQFHLNLSFPRKQRYVADGHIWACAFSR 347
 QY 642 NYGOKIKK-KNLCKRNFMLHVSMDENLISIMSIDKAVTKLRMO-----OKLE 689
 DB 348 FYEELHRSYSLFPCWRLFLIKLMHGLVDSATINNNTILLECRNSDPTTNNNSVD 407
 QY 690 KGESASPAEITEONGTANGSEISKEALETDSVS 728
 DB 408 RPSDSMTNNNNIVDHPN-----DINNKNVNDKNNNS 439

RESULT 9

094CF5 PRELIMINARY; PRT; 440 AA.
 AC 094CF5;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Hypothetical 50.6 kDa protein.
 GN D14450W.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosida II; Brassicales; Brassicaceae; Arabidopsids.

OX NCBI_TaxID=3702;

RA YAMADA K., LIU S.X., SAKANO H., PHAM P.K., BANH J., CHUNG M.K.,
 RA GOLDSMITH A.D., LEE J.M., QUACH H.L., TANG C., TORIUMI M., YU G.,
 RA BOWSER L., CARNINCI P., CHEN H., CHEUK R., HAYASHIZAKI Y., ISHIDA J.,
 RA JONES T., KAMIYA A., KARLIN-NEUMANN G., KAWAI J., KIM C., KOESAMA E.,
 RA LAM B., LIN J., MEYERS M.C., MIRANDA M., NARUSAKA M., NGUYEN M.,
 RA PALM C.J., SAKURAI T., SATOU M., SEKI M., SHIN P., SOUTHWICK A.,
 RA SHINOZAKI K., DAVIS R.W., ECKER J.R., THEOLOGIS A.;
 RA "Full length cDNA of gene d14450w.";
 RA Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Banh J., Banno F., Chang E., Dale J.M., Goldsmith A.D.,
 RA Lee J.M., Yoder C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,
 RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Kim C., Koesama E., Lam B., Lin J., Meyers M.C., Miranda M.,
 RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
 RA Shin P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RA "Full length cDNA of gene d14450w.";
 RA Submitted (NOV-2001) to the EMBL/Genbank/DBJ databases.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA EMBL; AY034902; AAK59409.1;
 DR EMBL; AY063047; AAL34221.1;
 DR InterPro: IPR007087; Znf.C2H2.
 DR SMART: SM00355; Znf.C2H2.1;
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 1.
 DR Hypothetical protein.
 SQ SEQUENCE 440 AA; 50623 MW; 0E0AB2C5517BE6F CRC64;

Query Match 5.6%; Score 217; DB 10; Length 440;
 Best Local Similarity 21.2%; Pred. No. 4.8e-07;
 Matches 80; Conservative 62; Mismatches 140; Indels 96; Gaps 11;
 QY 413 RKEQDTNENRQKIRITQYQFLYNNNTQQTREARDLHCPCWCTLNCRLYSILKLKLSHS 472
 DB 58 RKRK-----SRSTGVVFNKDCNNTLQKTEVREDSCPCSMCGSFKGLQFHLNSSH 112
 QY 473 RFTFNYYHFKGARIDVSI-----NECYDG-----SYAGNPQDIIHROPGFARNG 518
 DB 113 LFEFEFLSEYQVNVSVKLSNFIIEEGSDDDKFEFSLCSKPRKRRRQGGNNNTRL 172
 QY 492 NECY-----DGSYA--GNPQDIIHROPGFARNGPVKRTPTTH--- 527
 DB 173 KVCFLPLDSPSLANGTENGITLLNDNGRGLGYPEATELAGQFEMTSNIP--PAIASSL 229
 QY 528 -----ILVCRPKRTKASMEFELESDGVEEQRTYSSGHNLFYHSDCLPLRPOEY--E 580
 DB 230 DAGAKVILTTEAVVPATRTKLSAERSEA--RSHLLQKQFYSHRVPALAEQVMSD 286
 QY 581 VDSDEKPEWLREKTIQIEEFSQVNGEKEVKMLNLMHYMKHGFADNOMNHACMLFEVE 640
 DB 287 RDSDEVDVADFEEDRMDDFDVYNDKDEQFHLNLSFPRKQRYVADGHIWACAFVS 346
 QY 641 ENTGOKI-IRKNLCRNFMHLVSMHDENLISIMSIDKAVTKLRMOOKLEKESASPAE 699
 DB 347 RFEYKEHLHRSYSLFPCWRLFLIKLMHGLVDSATINNNTILLECR-----NT 394
 QY 700 EITEONGTANGSEIS 717
 DB 395 SVTNNNNNSVDHPDSMT 412

RESULT 10

092NT9 PRELIMINARY; PRT; 692 AA.
 AC 092NT9;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE FERTILIZATION-Independent seed 2 protein.

GN F152.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
 OK NCBI_TaxID=3702;
 RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99093530; PubMed=9874812;
 RA Luo M., Billodeau P., Koltunow A., Dennis E.S., Peacock W.J.,
 Chaudhury A.,
 RT "Genes controlling fertilization-independent seed development in
 Arabidopsis thaliana."
 RT Proc. Natl. Acad. Sci. U.S.A. 96:296-301(1999).
 RL EMBL; AF096096; AAD09105.1; -
 DR EMBL; AF096095; AAD09104.1; -
 DR InterPro: IPR007087; Znf_C2H2.
 DR SMART; SM00355; Znf_C2H2.1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
 FT VARIANT 637 W->*.
 SQ SEQUENCE 692 AA; 77666 MW; 2A3F61B961676497 CRC64;
 Query Match 4.9%; Score 190; DB 10; Length 692;
 Best Local Similarity 18.8%; Pred. No. 6.8e-05;
 Matches 126; Conservative 99; Mismatches 267; Indels 178; Gaps 24;
 DB 158 QLTFFGFHKNDKPSNSEQNSVTLVYLAVY-----CHKRRDYSCPIR 204
 74 QLSPLTFCSKRNQRORDSNVYKKNVLMELDLDLPGTENDSTHVNDNVSSPPR 133
 205 -----OVTGKQVPLIPDLNQTGPNF--PSLAVSNE----- 236
 134 AHSEKISDILTTTOLAIESSEPKVPHVND--GNVSSPPRAHSAEKNESTHVND 190
 237 -FEPSNSHWKSYSLF---RVTRPGREFNGMNGENIDVNEE--LPARRRNR 288
 191 VSSPPRAHSAEKNESTHVNDNISPPKAH---SSKKNSTHMDDEVSPPTRSSK 245
 289 EDGEKTFVAAQMTVPDK-----NRRQLDGEYVAQMECEPISKRAWTETLDG 340
 246 ETSDDLTTTQPAIVEPSEPKVRRGSRKQLYAKRY-----KARETQPAIAES 292
 341 KRLPPE-----ETFSQGP-----TLQFTLRMTGE-----TNDKSTAPIAKPIA 378
 293 SEPKVLHVNDENVSSPPEAHSALEKASDILTTTQPAIAESSEPKVPHVNDENVSSPTRAH 352
 379 TRNSSLHQ--ENKPGSVKPTQTIIVKESLTTDLQTRKEKDP-----NENRQKRIIFY 430
 353 SKKNSTKRVNDNVSPPTRSSKTSIDLTTTQPAIAESSEPKVRRHVNDNVSSPTRAH 412
 431 QPLVNNNTROQTEARDLHCPCWCTLNCRLYSILKHLKLCCHSRFIPNYVYHPGARIIVS 490
 413 SSKKNKSTRKND--DNIPSPPTRSSKTSNLTTRTQ-----PAIASEPK 456
 491 INECYDGSVAGNP-----ODIHRQGF-----FSRNGVPKRTPIIH 527
 457 VPHVNDKVSSTPRAHSSKKNSTKHKKDNASLPTTRSSKTSIDLATTQPAKAP--- 513
 528 ILVCPKRTKASMSFLESEDEGEVQQRITYSSGHNRLYHSDTCLPLRQEM--EYDSED 585
 514 ---SEPKTVRSRRLEAERCEAKRLEBK--GROFVHSQTMQPIFEQVYNSNDSN 567
 586 EKDPMLAREKTTTQILEFSDVNEGEKEVKKLNLNLMHMGFTADNOMNACHMLFVE--- 641
 568 ETDDVALDISERLERLVGSKSEKRYMYLMNIFVRKQRYIADGIVPACCEFAALHKE 627
 642 -----NYGOKIIRKNCNFMPLHLVSMDFNLISIDKAVTKLEMOQKLEKGS 693
 628 EKNSSSFDMMWRMRIRIKLMNGLICATFHKCTTILSNSDEA-----GQFTSGSA 679
 694 ASPANEELTE 703
 680 ANANNQOSME 689

RESULT 11
 ID 0920P0 PRELIMINARY; PRT: 632 AA.
 AC 0920P0;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Fertilization-independent seed 2 protein.
 GN AT2G35670.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
 OK NCBI_TaxID=3702;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=Cv. Columbia;
 RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,
 RA Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.-I.,
 RA Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Niernan W.C.,
 RA Fraser C.M., Venter J.C.,
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC006068; AAD15448.2; -
 DR InterPro: IPR007087; Znf_C2H2.
 DR SMART; SM00355; Znf_C2H2.1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
 SQ SEQUENCE 632 AA; 71130 MW; DDFDE33DC6620AE8 CRC64;
 Query Match 4.6%; Score 180; DB 10; Length 632;
 Best Local Similarity 18.5%; Pred. No. 0.0003;
 Matches 120; Conservative 92; Mismatches 243; Indels 192; Gaps 22;
 DB 158 QLTFFGFHKNDKPSNSEQNSVTLVYLAVY-----CHKRRDYSCPIR 204
 74 QLSPLTFCSKRNQRORDSNVYKKNVLMELDLDLPGTENDSTHVNDNVSSPPR 133
 205 -----OVTGKQVPLIPDLNQTGPNF--PSLAVSNE----- 236
 134 AHSEKISDILTTTOLAIESSEPKVPHVND--GNVSSPPRAHSAEKNESTHVND 190
 237 -FEPSNSHWKSYSLF---RVTRPGREFNGMNGENIDVNEE--LPARRRNR 288
 191 VSSPPRAHSAEKNESTHVNDNISPPKAH---SSKKNSTHMDDEVSPPTRSSK 245
 289 EDGEKTFVAAQMTVPDK-----NRRQLDGEYVAQMECEPISKRAWTETLDG 340
 246 ETSDDLTTTQPAIVEPSEPKVRRVRSRKLQYAKRY-----KARETQPAIAES 292
 341 KRLPPE-----ETFSQGP-----TLQFTLRMTGE-----TNDKSTAPIAKPIA 378
 293 SEPKVLHVNDENVSSPPEAHSALEKASDILTTTQPAIAESSEPKVPHVNDENVSSPTRAH 352
 379 TRNSSLHQ--ENKPGSVKPTQTIIVKESLTTDLQTRKEKDP-----NENRQKRIIFY 430
 353 SKKNSTKRVNDNVSPPTRSSKTSIDLTTTQPAIAESSEPKVRRHVNDNVSSPTRAH 412
 431 QPLVNNNTROQTEARDLHCPCWCTLNCRLYSILKHLKLCCHSRFIPNYVYHPGARIIVS 490
 413 SSKKNKSTRKND--DNIPSPPTRSSKTSNLT----- 443
 491 INECYDGSVAGNPQDILHRQGFASRNGVPKRTPIIHILVCPKRTKASMSFLESEDE 550
 444 -----ATTQPAKAP-----SEPKTVRSRRLEAERCE 473
 551 VEQQRITYSSGHNRLYHSDTCLPLRQEM--EYDSEDEKDPMLAREKTTTQILEFSDVE 608
 474 AKRLEBK--GROFVHSQTMQPIFEQVYNSNDSNEDVDYALDISERLERLVGSK 530
 609 GEKSVAKLNLNLMHMGFTADNOMNACHMLFVE-----NYGOKIIRKNCNFM 656
 531 EERRYMYLMNIFVRKQRYIADGIVPACCEFAALHEEKNSSSFDMMWRMRIRIKLMNNG 590

QY 657 MLHLVSHMDFNLISIMSDKAVTKLRMOQKLEKESASPANEITE 703
 DB 591 LICKATPHKCTILLNSDEA-----GQFTSGSANANNQSM 629

RESULT 12

Q8AMB3 PRELIMINARY; PRT; 570 AA.
 AC Q8AMB3;
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Kεatalin gamma 2.
 OS Lampetra fluviatilis (River lamprey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
 OC Petromyzontiformes; Petromyzontidae; Lampetra.
 ON NCBI_Taxid=7748;

RP SEQUENCE FROM N.A.
 RA Schultess J., Schaffeld M., Markl J.;
 RT "Unusual intermediate filament proteins from Lampetra fluviatilis.";
 RL Submitted (AUG-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AJ308111; CAC87101.1; -
 SQ SEQUENCE 570 AA; 58991 MW; FFAE0FFD3DDC872 CRC64;

Query Match 4.48; Score 171; DB 13; Length 570;
 Best Local Similarity 21.88; Pred. No. 0.0011;
 Matches 106; Conservative 61; Mismatches 162; Indels 158; Gaps 21;

QY 7 GGGGGGSGPSAGSGGGG-----GGSAAVAATAATSGKSGSGSGSGGSY 52
 DB 47 GGGGGGGGSLGGGGGFGALSTYGRGYGAGGGGAGLGRGLAQFGGGGGGGGGG 106
 QY 53 SASSS-----SSAAAAAGAAVLPRKKPKMEHYQADHELFLQAFKPTQYRF 99
 DB 107 GSSASFVGVGRSMASGGLHGAALAGGRSLP----- 137
 QY 100 LRRRLNLAIPFLHRTLLTYMSHRNSRINIKRRTKFEVDDMLSKVKMGESHSLSAHLQ 159
 DB 138 -----PL-----VTREGQRTVLRSLINDRFAGYIEKQRL--QEEVQLAHLK- 178
 QY 160 TETGFHKNKPKPNSP-----NEONSVTLTEVLVLYKCHKKR 197
 DB 179 SLTG--GVSAVPDSTELAGDPTARFVLEAFDALVLANQLEILDNLRAFAQEKAK 236
 QY 198 DVSCPIROVPTG--KKOYPLIPDLNQTAKGNF-----PSLAVSSNEFEPSNSHVSYS 249
 DB 237 -----YEFVGVAYQL-----ETDIGNMRKDLGATELKY--ELESRYSLMLELE 280
 QY 250 LLPRV-----TRGGRFNGMNGETNENT-----DVNEELPARKRRKREDCEK 293
 DB 281 FLKPAQAEEELKLGAKPKDKVDSGALIDVDSRSIDLGLMDMAKEVEALAKRRADADA 340
 QY 294 TFAVNAOTVFDRNRRLQLDGEYEVAQMECEPISKRRATWETILDKRLPRPFTFSQRP 353
 DB 341 YFQNAQTIEDH--KASSETESATSEITEVQASMOELOLE--LQG-----LSKNG 389
 QY 354 TLOFTL-----RWGTETNDRSTADIANPLATRNSESLHQBKPGSVKPTQTIAVESLTT 408
 DB 390 QLNHMLAEVGRYFEINTMQSR-VADMDGTLHSSKTELQRO--LVAFQDLIDKILNDT 446
 QY 409 DLQTRKE 415
 DB 447 EIATYKO 453

RESULT 13
 Q9MAK1 PRELIMINARY; PRT; 1745 AA.
 AC Q9MAK1;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)

DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 DE F27F5.16.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 ON NCBI_Taxid=3702;

RP SEQUENCE FROM N.A.
 RA Chao O., Brooks S., Buehler E., Johnson-Hopson C., Khan S., Kim C.,
 RA Shinn P., Altafi H., Bet O., Chin C., Chou J., Choi E., Conn L.,
 RA Conway A., Gonzales A., Hansen N., Hong B., Koo T., Lam B., Lee J.,
 RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
 RA Palm C., Pham P., Sakano H., Schwartz J., Souhailick A., Thaverl A.,
 RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
 RA Ecker J.R.;

RT "Genomic sequence for Arabidopsis thaliana BAC F27F5 from chromosome
 I.";
 RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AC007915; AF69170.1; -
 DR InterPro: IPR000087; Collagen.
 DR InterPro: IPR003653; SUMO protease.
 DR Pfam: PF02902; Peptidase C48; 1.
 DR PROSITE, PS50600; DUF-PROTEASE; 1.

SQ SEQUENCE 1745 AA; 194607 MW; 26BD44B87389E2F9 CRC64;

Query Match 4.48; Score 169.5; DB 10; Length 1745;
 Best Local Similarity 20.38; Pred. No. 0.0063;
 Matches 124; Conservative 74; Mismatches 218; Indels 195; Gaps 24;

QY 8 GGGGGGSGPSAGSGGGGSGSAAVAATAATSGKSGSGSGSGGSYSSSSSSAAAAAGAA 67
 DB 423 GGPSCGDEGAPSGSGDDEG-----PSCGDEGPGSGDGCPCGPGADGEGPGNG-- 472
 QY 68 VLPVKPKMEHYQADHELFLQAFKPTQYR--FLTRNLI-----APFLHRTL---- 115
 DB 473 -----ADGEVDEAFDKYAGELRSFKRDTOLDKFFEDLKFVREIRTAAL 519
 QY 116 -----TYMSHRNSTN--IKRTFKYDMLSKVKMGESHSLSAHLQPTTF 164
 DB 520 OSODRKEGPRSSKSDSPSEKVERVYTKVEKAVKREKANKREKAAEDVQRSV---- 575
 QY 165 FHKDKPS---PNSENQNSVTLTEVLVLYKCHKKRKVSCPIROV-----TGKKQVP 214
 DB 576 -KSTKPRKRVPRSSRLNNTP-----KKAATGSLPVEEVPONDGEGEYSAQTV 624
 QY 215 LIPLDNTKGNFSLAVSSNEFEPSNSHVKYSLSLF-----RYTRPGRE 261
 DB 625 SDYMENTIYLG--ASSTEEEGGNSLEEDSSKLHONVSVMEDNDVDRVPSHP 680
 QY 262 FNGMNGETNENT-----DVN-----EELPARKRRRE 289
 DB 681 EHGIPDGNFQLPDLILSDPALEKMSDVSSSHODVOKGLGNGDEEVAVESQOLE 740
 QY 290 DGEITFAQNTVF-----DKNRLQLDGEYEVAQMECEPISKRRATWETILDKRL 343
 DB 741 EPOSPKVKNOEFVEYTEGNPSAETICGEGEALDE-----DKSPVVADELDTAL 794
 QY 344 PPEFTSQCPITLQTLMTGETNDS--TAPIAKPLATRNSESLHQBKPGSVKPTQTIA 401
 DB 795 PGFVS-----PTVYDFEYENANLSPSSPTVYVSKVL--TQLKDILANVS---KIPKVA 846
 QY 402 VKESLTTDL-----QTRREKD-----TPNENRQRLRIYQ 431
 DB 847 VPBEVLTLQKDDVLEKEVSEKVAIPBETGKEDDVVEAGVSKTEAIVAPVNEKEKKSRR 906
 QY 432 FLVNNNTROQ-----TEARDLHCPCWCTLN--CKRLSLKHLK 468
 DB 907 VTFSDDTKNDKDKDPDVTVEERREAGPEVLVKAGDSLNLNLEKVEAKFORLCTITG 966
 QY 469 LCHSRFIENVY 479
 DB 967 RCHTLFGFRKY 977

RESULT 14

062007 PRELIMINARY: PRT: 620 AA.
 AC 062007: TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)
 DE Intermediate filament protein D1.
 OS Branchiostoma lanceolatum (Common lancelet) (Amphioxus).
 OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
 OC Branchiostoma.
 NCBI_TaxID=7740;
 RN NCB1
 RP MEDLINE-96267218; PubMed-9602172;
 RX MEDLINE-96267218; PubMed-9602172;
 RA Riemer D., Karabinos A., Weber K.:
 RT "Analysis of eight cDNAs and six genes for intermediate filament (IF)
 RT proteins in the cephalochordate Branchiostoma reveals differences in
 RT the IF multigene families of lower chordates and the vertebrates.";
 RL Gene 211:361-373(1998).
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 DR EMBL; AJ223579; CA11446.1; -.
 DR InterPro; IPR001664; IF.
 DR InterPro; IPR002957; Keratin_I.
 DR InterPro; IPR003054; Keratin_II.
 DR InterPro; IPR000772; Rictin_B_lectin.
 DR Pfam; PF00038; filament_1.
 DR Pfam; PF00652; Rictin_B_lectin; 3.
 DR PRINTS; PR01248; TYPE1KERATIN.
 DR PRINTS; PR01276; TYPE2KERATIN.
 DR SMART; SM00458; RICTIN; 1.
 DR PROSITE; PS00226; IF; 1.
 DR Collected coll: Intermediate filament.
 SQ SEQUENCE 620 AA: 67101 MW; 835AB189BF80A91B CRC64;

Query Match 4.3%; Score 166.5; DB 5; Length 620;
 Best Local Similarity 23.0%; Pred. No. 0.0026;
 Matches 99; Conservative 61; Mismatches 160; Indels 111; Gaps 19;

```

0Y 7 GGGGGGGSPAGSGGGG-----FGGS-----AAVAAATAGSGSGGS 45
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 31 GGGGGGGGSSSSSSGGGGGRASFGGSGGCGCGGRTSMRSMTNRSGAGGGRGGG 90
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
0Y 46 GGGGGSYASSSSSAAAAAAGAVLPVKPKMEHVOADHEL-----FLQAFKPTQYRF 99
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 91 -GGGGGRGYGSSMTAEQASQALVALGVYRVERTDKDELVLNDRFAFIETK----VAF 145
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
0Y 100 LTRNLIAPIFLHRTLYTWSHRNSRNTI---KRTFKVDDMLSKVEKMG--EQESHSL 153
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 146 LENQNR---KLEMLKRWQKSGSPDLGAMWEALRQLIEVYNTREGSLAEARGL 201
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
0Y 154 SAHQUTLTGFFHKNDKSPNSNENQNTL-----EVLVAVCKKKRVDSCPIRQVPT 208
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 202 SGEVELKTRY--DDEVGTRDGLLEIKIKIRADEFDESLTRVDLEARD-----S 249
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
0Y 209 GKKQVPLIPDLNQTGKGNFSLAVSSN--EPEPNSHWKSYSLFRV-----TRPCR 259
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 250 IKSELEPLKEYAAMIEALNSQILDSTMLELPSAGPVDLDSCLAEVAAQYEQLTRMR 309
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
0Y 260 REFNGMT-----NGETNENI-----DVNEELPARRKRR----- 288
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 310 AEAESWYATKFEDLQNSGKNNNDLADARSELSTKYONQIARLOSEIESKNNRQLEGOL 369
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
0Y 289 ---EDGEKTEVA--OMTVFDKRRRLQOLDGEVVAQMEHE-----CPISKRRATWETL 338
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 370 KNVESGSKLADKAETLEALFAELQRLGELSKQREYQELHNVKMLDVEIAAYRKL 429
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
0Y 339 GGR--RLPPE 347
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 430 EGESRLHGF 440
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

RESULT 15

081519 PRELIMINARY: PRT: 2359 AA.
 AC 081519: TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)
 DE Erythrocyte membrane protein 1 (PfEMP1).
 GN PF119600.
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 NCBI_TaxID=36329;
 RN NCB1
 RP SEQUENCE FROM N.A.
 RC STRAIN-3D7;
 RX MEDLINE-22255705; PubMed-12368864;
 RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
 RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
 RA Chan M.S., Nene V., Shalhou J., Sub B., Peterson J., Anguoli S.,
 RA Perera M., Allen J., Selengut J., Haft D., Mather M.W., Valdivia A.B.,
 RA Martin D.M.A., Fairhead A.H., Fraunholz M.J., Roos D.S., Kalpin S.A.,
 RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
 RA Fraser C.M., Barrell B.;
 RA "Genome sequence of the human malaria parasite Plasmodium
 RT falciparum.";
 RL Nature 419:498-511(2002).
 DR EMBL; AB014850; AAN36478.1; -.
 SQ SEQUENCE 2359 AA: 261601 MW; 29401D4450AD4DB5 CRC64;

Query Match 4.2%; Score 165; DB 5; Length 2359;
 Best Local Similarity 20.0%; Pred. No. 0.019; Mismatches 172; Indels 138; Gaps 18;
 Matches 94; Conservative 65;

```

0Y 7 GGGGGGGSPAGSGG-----GGFGSAVAAATAGSGSGGSGGSGYASSSSSAAA 62
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 631 GGGGGGGSGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 690
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
0Y 63 AAGAAVLPVKPKMEHVOADHELFLQAFKPTQYIFLTRLNIAPIFLHRTLYTWSHRN 122
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 691 GSGS-----KDSQKLYEHW-----CYRHDEVQRYOGG 720
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
0Y 123 SRTNIRKTFKYDDML-----KYKKMGEOESHLSAHLQLTFTGFF----- 165
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 721 EEDD-EDVEEEDDLKASAGCLCLPVPKKESEAKSQNDPEQFTFDFTFYVAHML 779
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
0Y 166 -----HKNDKSPNSNENQNTLLEVLVAVCKKKR--DVSCPIRQVPTGKQVPL 216
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 780 KDSIHWRTKLDKCIINTNGKQK-----CRKGGKGDGCDCLKRVNNEKTEMGKI 829
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
0Y 217 PDLNQTGKGNFSLAVSSNEFE--PSNSHWKSYSLFRVTRPG--RRENGMTNGET 270
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 830 KDHFKTQEG-----LDNENGCIPESTRFGFTHDVVD-----GEFLELOFENNEEN 877
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
0Y 271 NENIDVNEEL-----PARRKRRNREDEKTFVAQVTDKNNRQLQLDGEVVAQMEHE 323
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 878 SKNSLDSSEIKHLKLNLDIILKKKNQAGAGA-TGKATTND-----ELIDYESIADCCQ 931
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
0Y 324 ECPISKRRATWETLIDGKRLPPEPFSQPTLOFTLRMTGETNDKSTAPIARPLAT---- 379
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 932 KCOETQPOQVDRDLARSATGPRDPS-----PAAPTAGVESAAYD 973
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
0Y 380 -----RNSSEIHOENKRGSKVPTQIYAKESLITDLOTRKKEKIPNE 421
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 974 DEDDDNEEEEEEDEDEKEEVGGSDVVEKEE-----KEEKEPDEE 1015
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

Search completed: September 2, 2003, 11:04:15
 Job time : 60.1833 secs

THIS PAGE BLANK (USPTO)

PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142277.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.

PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 5.9%; Score 230; DB 21; Length 498;
Best Local Similarity 22.5%; Pred. No. 1.3e-09;
Matches 85; Conservative 62; Mismatches 161; Indels 70; Gaps 12;

QY 413 RKEKDTPENRQKRLRYOFLYNNNTROOTEANDDLHCPCWCTLNCRLYSLHLKIKCHS 472
Db 123 RKRK-----SRSTGMVVFYKDCNNNTLOKTEVEDSCSPCSMLCGSFKGLQFHLNBSHD 177
QY 473 RFFFNYY-----HEKGRIDV-SINEXY----- 495
Db 178 LFEFEFFYISVNPFTMECSSKPKRRRGGRNNTRLKVGFLPLDPSLTNGENGITLL 237
QY 496 -DGSYA-GNPDIHROGFAFSRNGPVKRPPIYH-----ILVCPKRTKASMSPELE 545

ID AAY53933 standard; Protein: 604 AA.
 AC AAY53933;
 DT 13-MAR-2000 (first entry)
 XX
 XX An Os-MPC1 protein having flowering regulating activity.
 DE
 KW flowering regulating activity; MPC1; flowering; germination;
 KW super early flowering mutation; altered flowering time;
 KW flowering regulating gene; food crop; vegetable; flowering inhibition;
 KM productivity.
 OS
 OS Oryza sativa.
 FH Key Location/Qualifiers
 FT Domain 310..335
 FT /note="zinc finger domain"
 FT
 PN EP967278-A2.
 PD 29-DEC-1999.
 XX
 XX 28-JUN-1999; 99EP-0305077.
 PF
 PR 26-JUN-1998; 98JP-0180065.
 PR 24-JUN-1999; 99JP-0179043.
 XX
 XX (MITA) MITSUI CHEM INC.
 PI Yoshida N, Kato Y, Takahashi S, Yanai Y, Hiratsuka J, Miwa T;
 DR WPI; 2000-064612/06.
 DR N-PSDB; AA236953.
 XX
 PT Novel DNA used to produce transgenic plants with altered floral
 PT regulation which can have increased crop yields -
 XX
 PS Claim 3; Page 36-39; 53pp; English.
 XX
 CC The present sequence represents a protein having a flowering regulating
 CC activity, which is designated Os-MPC1. An Arabidopsis MPC1 is also
 CC disclosed in the specification. The rice and Arabidopsis cDNAs show
 CC significant homology with each other. A naturally occurring mutation
 CC of the MPC1 gene eliminates normal flowering regulating ability of
 CC plants, and leads to flowering immediately after germination (super
 CC early flowering mutation). The MPC1 polynucleotide sequence can be
 CC used to produce plants with altered flowering times in comparison
 CC with wild type plants, by enhancing or inhibiting the expression of
 CC the flowering regulating gene. Antisense polynucleotides can be used
 CC to reproduce the effects of the mutated MPC1 gene. This alteration
 CC can be used to increase the yield of food crops. Flowering inhibition
 CC of vegetables increases their productivity.
 CC
 SO Sequence 604 AA;
 Query Match 6.3%; Score 243.5; DB 21; Length 604;
 Best Local Similarity 19.1%; Pred. No. 1.4e-10;
 Matches 134; Conservative 97; Mismatches 249; Indels 223; Gaps 23;
 QY 77 EHVOADHLEFLQAFKPTQIYRFLRNLAPILFLRLTYSH--RNSRTNI----- 127
 DB 14 EQLAAEESPAL--YCPPELVNINIORRSIKNPALFQRCLLYIHAARKRRSLTISLGG 71
 QY 128 KRRTFVDDMLSKVEKMGESHSLSAH-----LQLTGEGFHKKDKPS-----P 173
 DB 72 TKRELRAONIFPLLYLAPLPTNNVSLGHSPIYRSLCLTSFHEFGKNDTEAFVAP 131
 QY 174 NSEN--EONSVTLEVLVAVCHKKRK--DVSCTPIQVPTGKKQ-----VPLIPD 218
 DB 132 DVKNLATSRACSLNITLIS--CGRAEQTFFDNNCSGHNHVGSTLQKLEGCFMCKIFI--D 188
 QY 219 LNOTRGNPPLAVSSNEPEPSNHWKYSILFLVTRPGRREFGNGINGETNENIDVNE 278

DB 189 LLAASLGNCSLSL-----GHTVEMSSIVE 213
 QY 279 ELPA--RRKRNREGSKTPVAQMTVPDKNRLQLDGVEYVAMQEMEBCISKRAWTET 336
 DB 214 MTPSLEPKFLEDSCLTFCOSKVADATGSGFQLOVASTSAGAKDNSEPSY----- 285
 QY 337 ILDGKRLPPEETFSOGPTLQFTLRWTGETNDKSTABIANPLATRNSESLHQENKPGSVXP 396
 DB 266 VYSNDVPP----- 274
 QY 397 TOTIANKESLITDLOTRKEKOTPNENRQKLRITPYQLYNNNTROGTARDLHCPMCTLN 456
 DB 275 -----SSLTHIIRLRSG-----NVLFNKYNNNTQKTEVEDEFCPLCLVP 316
 QY 457 CRRLVSLMKHLKCHSRPIFNYYHPKCARIDVSI--NECDGSAQNPQDIHQPFAPF 515
 DB 317 CGSEFKGLGCHLNSHDLFHEWISSECAVANSLTDSKRIPELLAEGVDPRHQTESYRS 376
 QY 516 RINGPVKRPIT-----HILVC--RPKRTKASMESEFLESEDEGEEOQRTYSSGH----- 561
 DB 377 RFKKRRVELSSDKIHVHPHIVDSGSPEDAQAG-----SEDDYVQRENGSSVAAIASVD 430
 QY 562 -----NELYHSPTCLPLRQEM--EV 581
 DB 431 PANSLSGSLSAFTVLQFGKTRKLSYERADPRNRLLQKQFPHSHRAQPMAMSKYFSDR 490
 QY 582 DSEDEKDPKRLKRTTQIEEFDVNEGEKEVKMLNLHYMKHGFADNOMNACHLFE 641
 DB 491 DSEDEVDLADIEDRMDDFDVYTKDELHMHMNSFVRKORVLADGHITPCEAFSQ 550
 QY 642 NYGQKIIKKNL--CRNFMHLVSMHDFNLISISIDKAVTKL 681
 DB 551 FPGQELVQNPALMCWRFF--MYKIMNHSLDRAANACTIL 591
 RESULT 14
 AAY53932
 ID AAY53932 standard; Protein: 611 AA.
 XX
 AC AAY53932;
 DT 13-MAR-2000 (first entry)
 XX
 DE A MPC1 protein having flowering regulating activity.
 DE
 KW flowering regulating activity; MPC1; flowering; germination;
 KW super early flowering mutation; altered flowering time;
 KW flowering regulating gene; food crop; vegetable; flowering inhibition;
 KM productivity.
 OS
 OS Arabidopsis thaliana.
 OS
 FH Key Location/Qualifiers
 FT Domain 306..327
 FT /note="zinc finger domain"
 FT
 PN EP967278-A2.
 PD 29-DEC-1999.
 XX
 XX 28-JUN-1999; 99EP-0305077.
 PF
 PR 26-JUN-1998; 98JP-0180065.
 PR 24-JUN-1999; 99JP-0179043.
 XX
 XX (MITA) MITSUI CHEM INC.
 PI Yoshida N, Kato Y, Takahashi S, Yanai Y, Hiratsuka J, Miwa T;
 DR WPI; 2000-064612/06.
 DR N-PSDB; AA236947, AA236948.
 XX

XX Venter JC, Adams M, Li PWD, Myers EN;
PI WPI: 2001-656860/75.
DR N-PSDB: ABL15738.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from *Drosophila* and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 41697; 21bp + Sequence Listing: English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB720722).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 955 AA;
Query Match 30.3%; Score 1177; DB 22; Length 955;
Best Local Similarity 34.8%; Pred. No. 3e-85;
Matches 281; Conservative 136; Mismatches 225; Indels 166; Gaps 26;
43 GSGCGG--GGSYSASSSSAAAAAALVPR--KPRMEHVQADHELFLQAFKPTQIYR 98
15 GSAMGCIIGLHGAPDASN---AGSTVPPTAEGQVXINGHQOELFLQAFKPTQIYR 70
99 FLTRNLIAPLEIHRITLYSHRNSRNRKPKRYVDMSKYMKGQESHSLSA-HL 157
71 YLRNRHETNPLELKRITLYMKRKRNNKRISFQVNSMEST---TKSEVSONYL 125
158 QLTFTGFHNRKDRKSPENSEO-----NSVLEVLVYVCHKRRDVSCLPIQV 206
126 HVIYDSLHEKLPARDNESGELLQEQLEAGSESVETITLYKTRSKRKRDSLQDEL 185
207 PTGKKQVPLPLDNLQTPGNGFPLAASSNEEP-SNSHMVKSYSILRVTTPGRREENG 265
186 LSKCSQIVNPP---KDVVGEHATISIPQTRMPGEOHTL--YKLIFRIK-----V 231
266 INGETNINIVNEELPARRRKRNREDGKTPVAQMTVDKNRRLQDLGGEVANAQMEEC 325
232 LSPSTCN--DENATPP-NKRSRPN-EKMGSELILEKSSGF-TTEGETEAMLQPLNST 286
326 PI---SKKRAWETILDGKRLP--PEPESQGPFLQFTLRMTGETNDKSTAPIAKPLAT 379
287 SIKFSFSKCKWETMPD-SYIPLSLTYDVYQOOSPLKFLHLT---SNQOLPEMISABELO 342
380 R-----NSSLSHOEN-----KRGVAKPTQTIIVAKESLTTDQTRKEKDPNE 421
343 RYVOHLDAVAEMNNNNNNNNNNNSGSLKNGSGGNGSTVC-----KTRP-- 386
422 NRQKLRIFYOLFVLYNNNTROQTEARDLHCPMCTLNCNKRLKSLKHLKLSRFLFENYVH 481
387 --EHIQIYVNFMSNNRROQTEYTOELNCPMGGLDLRLALLHLKCHAREFTTQPA 444
482 PKGARIVSINECYDSYAGNPQDIIHROPGFASRN-GPVKRPPTIHLVCRPRTASM 540
445 GSGARIVTINDAIDGSGPYDLAPRSGSSFAITGPGVARTSVTSLMCRPRQKTC 504
541 SEFLSESDGEVEQRTSSGHNRLYFHSDTCLPLRQEMEYVSDERDPEMLRKRITQI 600
505 DEFLELDEDEISNORSYITGHNRLYHHTETCLPVHPKELDIDSGESDPLMLRQKTOI 564
601 EEFSDVNBGEKVKLNHLHYMKHGFADNQMNAACHLFEVNTQKIIKKLCAENFMHL 660
565 DEFSDVNBGEKELMNLHYMRHGFVGDCLPLACEMFLDAKGTETVRKMLYENFTLHM 624

OY 661 VSMHDFNLISIMSDKAVTKLRMOOKLEKSES----- 693
DB 625 CSLFDYGLIAETVYTVQKLOGLSLKYAAGELMOROREBOLKYWLDVGMKROEDPPT 684
OY 694 -----ASPANEITE-----EONGTANG 711
DB 685 LKSPQKRAPADQASSTASSTSGSSSSMPKRPMLKRGSAASSPGVSKGTENG 744
OY 712 FSEINSEKALEFDSVSGVSKSKOKL 739
DB 745 TNGSNS-----SSSNKRVAKKSDQPL 767
RESULT 11
ID AAU16402 standard; protein; 175 AA.
XX
AC AAU16402;
XX
DE 07-NOV-2001 (first entry)
XX
DE Human novel secreted protein, Seq ID 1355.
XX
KW Human; immunosuppressive; antiarthritic; antirheumatic;
KW cytoskeletal; cardiac; vasotropic; cerebroprotective; neotropic;
KW neuroprotective; antibacterial; virocidic; fungicide; ophthalmological;
KW vulnary; secreted protein; rheumatoid arthritis;
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW corneal infection; wound healing; epithelial cell proliferation;
KW skin ageing; food additive; preservative; antiproliferative.
OS Homo sapiens.
XX
PN WC20015322-A2.
XX
PD 02-AUG-2001.
XX
PE 17-JAN-2001; 2001WO-US01341.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.

XX haemostatic; antithrombotic.
 OS Homo sapiens.
 XX US2002132753-A1.
 XX 19-SEP-2002.
 XX 17-JAN-2001; 2001US-0764864.
 XX 31-JAN-2000; 2000US-179065P.
 PR 04-FEB-2000; 2000US-180628P.
 PR 28-JUN-2000; 2000US-214886P.
 PR 07-JUL-2000; 2000US-216647P.
 PR 07-JUL-2000; 2000US-216880P.
 PR 11-JUL-2000; 2000US-217487P.
 PR 11-JUL-2000; 2000US-217496P.
 PR 14-JUL-2000; 2000US-218290P.
 PR 26-JUL-2000; 2000US-220963P.
 PR 26-JUL-2000; 2000US-220964P.
 PR 14-AUG-2000; 2000US-224518P.
 PR 14-AUG-2000; 2000US-224519P.
 PR 14-AUG-2000; 2000US-225267P.
 PR 14-AUG-2000; 2000US-225268P.
 PR 14-AUG-2000; 2000US-225270P.
 PR 14-AUG-2000; 2000US-225447P.
 PR 14-AUG-2000; 2000US-225757P.
 PR 14-AUG-2000; 2000US-225758P.
 PR 22-AUG-2000; 2000US-225758P.
 PR 30-AUG-2000; 2000US-228924P.
 PR 01-SEP-2000; 2000US-229287P.
 PR 01-SEP-2000; 2000US-229343P.
 PR 01-SEP-2000; 2000US-229344P.
 PR 01-SEP-2000; 2000US-229345P.
 PR 05-SEP-2000; 2000US-229509P.
 PR 05-SEP-2000; 2000US-229513P.
 PR 08-SEP-2000; 2000US-231413P.
 PR 21-SEP-2000; 2000US-234223P.
 PR 21-SEP-2000; 2000US-234274P.
 PR 25-SEP-2000; 2000US-234997P.
 PR 27-SEP-2000; 2000US-235834P.
 PR 29-SEP-2000; 2000US-236327P.
 PR 29-SEP-2000; 2000US-236367P.
 PR 29-SEP-2000; 2000US-236368P.
 PR 29-SEP-2000; 2000US-236369P.
 PR 29-SEP-2000; 2000US-236370P.
 PR 02-OCT-2000; 2000US-236802P.
 PR 02-OCT-2000; 2000US-237037P.
 PR 02-OCT-2000; 2000US-237038P.
 PR 02-OCT-2000; 2000US-237039P.
 PR 02-OCT-2000; 2000US-237040P.
 PR 13-OCT-2000; 2000US-239935P.
 PR 20-OCT-2000; 2000US-240960P.
 PR 20-OCT-2000; 2000US-241785P.
 PR 20-OCT-2000; 2000US-241809P.
 PR 01-NOV-2000; 2000US-244617P.
 PR 17-NOV-2000; 2000US-249239P.
 PR 08-DEC-2000; 2000US-251856P.
 PR 08-DEC-2000; 2000US-251858P.
 PR 08-DEC-2000; 2000US-251869P.
 XX (ROSE/) ROSEN C A.
 PA (RUBEN/) RUBEN S M.
 PA (BARA/) BARASH S C.
 XX
 PI Rosen CA, Ruben SM, Barash SC;
 XX
 DR WPI, 2003-147444/14.
 DR N-PSDB; ABX73744.
 XX
 PT New polypeptides and nucleic acids, useful in gene therapy for
 PT treating, inhibiting or preventing e.g. neural, immune system,
 PT muscular, respiratory, reproductive, gastrointestinal, pulmonary,

PT cardiovascular or renal disorders -
 XX
 PS Claim 11; SEQ ID NO 1369; 402pp; English.
 XX
 CC The invention relates to human novel polypeptides and their associated
 CC polynucleotides. The polypeptides and polynucleotides are useful in gene
 CC therapy for treating, inhibiting or preventing neural disorders, immune
 CC system disorders (e.g. systemic lupus erythematosus, rheumatoid
 CC arthritis and multiple sclerosis), muscular disorders, respiratory
 CC diseases (e.g. nasal vestibulitis, nasal polyps and sinusitis),
 CC reproductive disorders, gastrointestinal disorders, pulmonary disorders,
 CC cardiovascular disorders (e.g. congenital heart defects, Ebstein's
 CC anomaly and hypoplastic left heart syndrome), renal disorders (e.g. acute
 CC kidney failure and end-stage renal disease), hyperproliferative disorders
 CC (e.g. Hodgkin's disease and leukemia), inflammatory diseases (e.g.
 CC septic shock, bursitis and appendicitis), allergic reactions and
 CC conditions (e.g. asthma), blood related disorders (e.g. thrombosis,
 CC atherosclerosis and myocardial infarction) and cancerous diseases.
 CC Sequences AB054914-AB055699 and AB055748 represent human novel
 CC polypeptides of the invention.
 XX
 SQ Sequence 292 AA;
 XX
 QY Query Match 37.4%; Score 1453; DB 24; Length 292;
 Db Best Local Similarity 96.8%; Pred. No. 3,5e-108;
 Matches 275; Conservative 1; Mismatches 8; Indels 0; Gaps 0;
 QY 196 RKDVSCPIRVPGKQVPLIPDLNQTGKGNFSLAVSSNEFEPSNSHWKSYSLFRYT 255
 Db 1 KHDVSCPIRVPGKQVPLIPDLNQTGKGNFSLAVSSNEFEPSNSHWKSYSLFRYT 60
 QY 256 RGRREFNGINGNETNENIDVNEELPARRKRNEDEKTFVQMTVEFKNRRLQLLDGEY 315
 Db 61 RGRREFNGINGNETNENIDVNEELPARRKRNEDEKTFVQMTVEFKNRRLQLLDGEY 120
 QY 316 EVAMQEMECPISKKATWETIIDGRLLPPEFEFSGPILOFLRTGETNDKSTAPIAK 375
 Db 121 XVAMQEMECPISKKATWETIIDGRLLPPEFEFSGPILOFLRTGETNDKSTAPIAK 180
 QY 376 PLATRNSESLHONKPGSVKPTIIVAKESLFTDLOTRKEDTPNENRKLRIFYOLYN 435
 Db 181 PLATRNSESLHONKPGSVKPTIIVAKESLFTDLOTRKEDTPNENRKLRIFYOLYN 240
 QY 436 NNTROOTEARDDLHCPWCTLNCRKLYSLKHLKLCRSREIFNVY 479
 Db 241 NNTROOTEARDDLHCPWCTLNCRKLYSLKHLKLCRSREIFNVY 284
 RESULT 10
 ABB71635
 ID ABB71635 standard; Protein; 955 AA.
 XX
 AC ABB71635;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 41697.
 XX
 KW Drosophila: developmental biology; cell signalling; insecticide;
 KM pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.

PT cardiovascular or renal disorders -
XX
PS Claim 11; SEQ ID NO 911: 402pp; English.
XX
CC The invention relates to human novel polypeptides and their associated
CC polynucleotides. The polypeptides and polynucleotides are useful in gene
CC therapy for treating, inhibiting or preventing neural disorders, immune
CC system disorders (e.g. systemic lupus erythematosus, rheumatoid
CC arthritis and multiple sclerosis), muscular disorders, respiratory
CC diseases (e.g. nasal vestibulitis, nasal polyps and sinusitis),
CC reproductive disorders, gastrointestinal disorders, pulmonary disorders,
CC cardiovascular disorders (e.g. congenital heart defects, Epstein's
CC anomaly and hypoplastic left heart syndrome), renal disorders (e.g. acute
CC kidney failure and end-stage renal disease), hyperproliferative disorders
CC (e.g. Hodgkin's disease and leukaemia), inflammatory diseases (e.g. acute
CC septic shock, bursts and appendicitis), allergic reactions and
CC conditions (e.g. asthma), blood related disorders (e.g. thrombosis,
CC atherosclerosis and myocardial infarction) and cancerous diseases.
CC Sequences AB054914-AB055699 and AB055748 represent human novel
CC polypeptides of the invention.
XX
SQ Sequence 289 AA;
Query Match 38.5%; Score 1497; DB 24; Length 289;
Best Local Similarity 99.6%; Pred. No. 1e-111; 0; Indels 0; Gaps 0;
Matches 289; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 456 NCRKLYSLNHLKCHSRFTFNTVYHPKARIDVINSINCYDGSVAGNPDIHQPGAFS 515
DB 6 DCRKLYSLNHLKCHSRFTFNTVYHPKARIDVINSINCYDGSVAGNPDIHQPGAFS 65
QY 516 RNPVVRPTPTHTLVCPRKTKASMSFLESDGEVQOQRTYSSGNHRLTFHSPTCLPLR 575
DB 66 RNPVVRPTPTHTLVCPRKTKASMSFLESDGEVQOQRTYSSGNHRLTFHSPTCLPLR 125
QY 576 PQMEVDSDEKPEMLREKTIQIEEFSDVNEGKEVMTLMNHLVYKRGFIADNOMNHA 635
DB 126 PQMEVDSDEKPEMLREKTIQIEEFSDVNEGKEVMTLMNHLVYKRGFIADNOMNHA 165
QY 636 CMLEFVNYGQIKIKKNCRNFMHLVSMHDFNISIMSIDKAVYTKLEMOQKLEKESAS 695
DB 186 CMLEFVNYGQIKIKKNCRNFMHLVSMHDFNISIMSIDKAVYTKLEMOQKLEKESAS 245
QY 696 PANEETTERONGTANGSEINSEKKALETDSVSGVSKOSKOKL 739
DB 246 PANEETTERONGTANGSEINSEKKALETDSVSGVSKOSKOKL 289
RESULT 8
AAU16416
ID AAU16416 standard; Protein; 292 AA.
XX
AC AAU16416;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human novel secreted protein, Seq ID 1369.
XX
XX Human; immunosuppressive; antiarthritic; antirheumatic;
KW cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW vulnerary; secreted protein; rheumatoid arthritis;
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW corneal infection; wound healing; epithelial cell proliferation;
KW skin ageing; food additive; preservative; antiproliferative.
XX
OS Homo sapiens.
XX
PN MO200155322-A2.
XX
PD 02-AUG-2001.

XX
PF 17-JAN-2001; 2001WO-US01341.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 26-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236337.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.

XX New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives -

PS Claim 11; SEQ ID No 911: 980pp; English.

XX
CC The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility
CC to a pathological condition. Antibodies to the proteins can also
CC be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiodenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection, and many other
CC disorders listed in the specification. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence represents a novel secreted protein of the invention.

Query Match 38.5%; Score 1497; DB 22; Length 289;

Best Local Similarity 99.6%; Pred. No. 1e-111;

Matches 283; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 456 NCRRLYSILKHLKLCCHSFINTYHYHKGARIDVSYNECTGYAGNPQDIHROPGEAFS 515
DB 6 DCRRLYSILKHLKLCCHSFINTYHYHKGARIDVSYNECTGYAGNPQDIHROPGEAFS 65
QY 516 RRGVYKRPPIHILVCRPKRTKASMESEFLESEDEVEOQRYSSGHNLYHSPCTPLR 575
DB 66 RRGVYKRPPIHILVCRPKRTKASMESEFLESEDEVEOQRYSSGHNLYHSPCTPLR 125
QY 576 PQEMEVDSEDEKDEMLKEKTIQIEESDVNEGEKEVYKLMNLVYKMGFIADNOMNHA 635
DB 126 PQEMEVDSEDEKDEMLKEKTIQIEESDVNEGEKEVYKLMNLVYKMGFIADNOMNHA 185
QY 636 CMLFVENYGGKIIKKNLCRNFMNLVSMHDPNLSISIDKAVYKLRMOOKLEKGEAS 695
DB 186 CMLFVENYGGKIIKKNLCRNFMNLVSMHDPNLSISIDKAVYKLRMOOKLEKGEAS 245
QY 696 PANEITEEONGTANGSEINSEKKEALETDSVSGVSKOSKOKL 739
DB 246 PANEITEEONGTANGSEINSEKKEALETDSVSGVSKOSKOKL 289

RESULT 7

ABU55027

ID ABU55027 standard; Protein: 289 AA.

AC ABU55027;

XX 18-MAR-2003 (first entry)

DE Human novel polypeptide #114.

KM Human; neural disorder; immune system disorder; renal disorder;
KM muscular disorder; respiratory disease; reproductive disorder;
KM gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KM hyperproliferative disorder; inflammatory disease; allergic reaction;
KM blood related disorder; cancer; immunosuppressive; antiinflammatory;
KM cardiovascular; nephrotoxic; cytostatic; antiallergic; thrombolytic;

KM haemostatic; antiarteriosclerotic.

XX Homo sapiens.

XX US2002132753-A1.

XX 19-SEP-2002.

XX 17-JAN-2001; 2001US-0764864.

PR 31-JAN-2000; 2000US-179065P.
PR 04-FEB-2000; 2000US-180628P.
PR 28-JUN-2000; 2000US-214866P.
PR 07-JUL-2000; 2000US-216647P.
PR 07-JUL-2000; 2000US-216800P.
PR 11-JUL-2000; 2000US-217487P.
PR 11-JUL-2000; 2000US-217496P.
PR 14-JUL-2000; 2000US-218290P.
PR 26-JUL-2000; 2000US-220963P.
PR 26-JUL-2000; 2000US-220964P.
PR 14-AUG-2000; 2000US-224518P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225267P.
PR 14-AUG-2000; 2000US-225268P.
PR 14-AUG-2000; 2000US-225270P.
PR 14-AUG-2000; 2000US-225447P.
PR 14-AUG-2000; 2000US-225757P.
PR 14-AUG-2000; 2000US-225758P.
PR 22-AUG-2000; 2000US-226868P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 08-SEP-2000; 2000US-231413P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234977P.
PR 27-SEP-2000; 2000US-235834P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239355P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241809P.
PR 01-NOV-2000; 2000US-244617P.
PR 17-NOV-2000; 2000US-249299P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.

XX (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.

PI Rosen CA, Ruben SM, Barash SC;

XX WPI: 2003-147444/14.

DR N-PSDB: ABX73286.

PT New polypeptides and nucleic acids, useful in gene therapy for
PT treating, inhibiting or preventing e.g. neural, immune system,
PT muscular, respiratory, reproductive, gastrointestinal, pulmonary,

XX 17-JAN-2001; 2001WO-US01341.
PF XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246509.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM.
PI WPI, 2001-488763/53.
XX N-PSDB; AAS25945.
DR

PF 17-JAN-2001; 2001US-0764864.
XX
PR 31-JAN-2000; 2000US-179065P.
PR 04-FEB-2000; 2000US-180628P.
PR 28-JUN-2000; 2000US-214886P.
PR 07-JUL-2000; 2000US-216647P.
PR 07-JUL-2000; 2000US-216880P.
PR 11-JUL-2000; 2000US-217487P.
PR 11-JUL-2000; 2000US-217496P.
PR 14-JUL-2000; 2000US-218290P.
PR 26-JUL-2000; 2000US-220963P.
PR 26-JUL-2000; 2000US-220964P.
PR 14-AUG-2000; 2000US-224518P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225267P.
PR 14-AUG-2000; 2000US-225268P.
PR 14-AUG-2000; 2000US-225270P.
PR 14-AUG-2000; 2000US-225447P.
PR 14-AUG-2000; 2000US-225757P.
PR 14-AUG-2000; 2000US-225758P.
PR 22-AUG-2000; 2000US-226868P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 08-SEP-2000; 2000US-231413P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234997P.
PR 27-SEP-2000; 2000US-235834P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239335P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241809P.
PR 01-NOV-2000; 2000US-244617P.
PR 17-NOV-2000; 2000US-249299P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
XX
XX (ROSE/) ROSEN C. A.
PA (RUBE/) RUBEN S. M.
PA (BARA/) BARASH S. C.
XX
PI Rosen CA, Ruben SM, Barash SC;
XX
XX MPI: 2003-147444/14.
DR N-PSDB; ABX73306.
XX
XX New polypeptides and nucleic acids, useful in gene therapy for
PT treating, inhibiting or preventing e.g. neural, immune system,
PT muscular, respiratory, reproductive, gastrointestinal, pulmonary,
PT cardiovascular or renal disorders -
XX
PS Claim 11; SEQ ID NO 931; 402pp; English.
XX
XX The invention relates to human novel polypeptides and their associated
CC polynucleotides. The polypeptides and polynucleotides are useful in gene
CC therapy for treating, inhibiting or preventing neural disorders, immune
CC system disorders (e.g. systemic lupus erythematosus, rheumatoid

CC arthritis and multiple sclerosis), muscular disorders, respiratory
CC diseases (e.g. nasal vestibulitis, nasal polyps and sinusitis),
CC reproductive disorders, gastrointestinal disorders, pulmonary disorders,
CC cardiovascular disorders (e.g. congenital heart defects, Epstein's
CC anomaly and hypoplastic left heart syndrome), renal disorders (e.g. acute
CC kidney failure and end-stage renal disease), hyperproliferative disorders
CC (e.g. Hodgkin's disease and leukemia), inflammatory diseases (e.g.
CC septic shock, buritis and appendicitis), allergic reactions and
CC conditions (e.g. asthma), blood related disorders (e.g. thrombosis,
CC atherosclerosis and myocardial infarction) and cancerous diseases.
CC Sequences AB054914-AB055699 and AB055748 represent human novel
CC polypeptides of the invention.
XX
SQ Sequence 388 AA;
Query Match 52.6%; Score 2043; DB 24; Length 388;
Best Local Similarity 99.5%; Pred. No. 2.2e-155;
Matches 386; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 352 GPTLQFTLRMTGETNDKSTAPIAKPLATRNSESLHQENKPGSVKPTQTIAVKESLPTTDQ 411
Db 1 GPTLQFTLRMTGETNDKSTAPIAKPLATRNSESLHQENKPGSVKPTQTIAVKESLPTTDQ 60
QY 412 TRKEKDPNENRQKRLIFQPLNNNTROOTEARDLHCPWCTLNCRLKLSLKHLKCH 471
Db 61 TRKEKDPNENRQKRLIFQPLNNNTROOTEARDLHCPWCTLNCRLKLSLKHLKCH 120
QY 472 SRFENFVYHPKGRIVDSINECDGSGYAGNPDIHQPGFARSRNPPVRRTPTHILVC 531
Db 121 SRFENFVYHPKGRIVDSINECDGSGYAGNPDIHQPGFARSRNPPVRRTPTHILVC 180
QY 532 RPKRTKASMSFELSEDEGEVQOQRTYSSGHNRLTFHSDTCLPLRPQMEVDSDEKDPEN 591
Db 181 RPKRTKASMSFELSEDEGEVQOQRTYSSGHNRLTFHSDTCLPLRPQMEVDSDEKDPEN 240
QY 592 LREKTIQIEFESVNGEKEKVMKLMVHKHGFIDNOMNACMLFVNTGOKITKN 651
Db 241 LREKTIQIEFESVNGEKEKVMKLMVHKHGFIDNOMNACMLFVNTGOKITKN 300
QY 652 LCRNFMHLVSMHDFNLSINSDIKAVTKLREMOQKLEKGSASPAHEITEBONGTANG 711
Db 301 LCRNFMHLVSMHDFNLSINSDIKAVTKLREMOQKLEKGSASPAHEITEBONGTANG 360
QY 712 FSEINSKEKALETDSVSGVSQSKQKRL 739
Db 361 FSEINSKEKALETDSVSGVSQSKQKRL 388
RESULT 6
AAU15958 standard; Protein: 289 AA.
ID AAU15958
XX
XX AAU15958;
AC
XX
DT 07-NOV-2001 (first entry)
XX
DE Human novel secreted protein, Seq ID 911.
XX
XX Human; immunosuppressive; antiarthritic; antirheumatic;
KW cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW vulnary; secreted protein; rheumatoid arthritis;
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW cerebrovascular disorder; cerebral ischemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW corneal infection; wound healing; epithelial cell proliferation;
KW skin aging; food additive; preservative; antiproliferative.
XX
XX Homo sapiens.
OS
XX
XX MO200155322-A2.
PM
XX 02-AUG-2001.
PD

Query Match	52.6%	Score 2043	DB 22	Length 388
Best Local Similarity	99.5%	Pred. No. 2.2e-155		
Matches 386	Conservative 0	Mismatches 2	Indels 0	Gaps 0
CC	include autoimmune diseases e.g. Rheumatoid arthritis,			
CC	hyperproliferative disorders e.g. neoplasms of the breast or liver,			
CC	cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders			
CC	e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.			
CC	Alzheimer's disease, infections caused by bacteria, viruses and fungi			
CC	and ocular disorders e.g. corneal infection, and many other			
CC	disorders listed in the specification. The polypeptides can also			
CC	be used to aid wound healing and epithelial cell proliferation, to			
CC	prevent skin aging due to sunburn, to maintain organs before			
CC	transplantation, for supporting cell culture of primary tissues, to			
CC	regenerate tissues and in chemotaxis. The polypeptides can also be used			
CC	as a food additive or preservative to increase or decrease storage			
CC	capabilities, fat content, lipid, protein, carbohydrate, vitamins,			
CC	minerals, cofactors and other nutritional components. The present			
CC	sequence represents a novel secreted protein of the invention.			
QY	352	GPPIQETLRMGNDKSTAPIANPLATRNSESILHOENKPGSVPTORTIAVRESLTITDQ	411	
DB	1	GPPIQETLRMGNDKSTAPIANPLATRNSESILHOENKPGSVPTORTIAVRESLTITDQ	60	
QY	412	TRKEKDPENENRQKRLIFQELVYNNNTQQOATEARDLHCPCCTLCRKLVSILKLKLC	471	
DB	61	TRKEKDPENENRQKRLIFQELVYNNNTQQOATEARDLHCPCCTLCRKLVSILKLKLC	120	
QY	472	SRFIFNYYHKGARIDVSIIECTDGSYAGNPQDIIHROPGAFSNGPVKRTPTTHILVC	531	
DB	121	SRFIFNYYHKGARIDVSIIECTDGSYAGNPQDIIHROPGAFSNGPVKRTPTTHILVC	180	
QY	532	RPKRTKASMSFELESDGEVECOORTYSSGSHRLTFHSPDCTPLRPOKEVSDSEKDEP	591	
DB	181	RPKRTKASMSFELESDGEVECOORTYSSGSHRLTFHSPDCTPLRPOKEVSDSEKDEP	240	
QY	592	LREKTIQIEEFSDVNEGEKEVMKLMNLHVKKHGFADQNQNHACMLFVENVGOKIIRKN	651	
DB	241	LREKTIQIEEFSDVNEGEKEVMKLMNLHVKKHGFADQNQNHACMLFVENVGOKIIRKN	300	
QY	652	LCRNFMHLVSMHDPNLIISIDSIDKATYKLEKMOQKLEKGSASAPANEITBEONGTANG	711	
DB	301	LCRNFMHLVSMHDPNLIISIDSIDKATYKLEKMOQKLEKGSASAPANEITBEONGTANG	360	
QY	712	FSEINSKEKALETDSVSGVSKOSKOKL 739		
DB	361	FSEINSKEKALETDSVSGVSKOSKOKL 388		
RESULT 5				
ABU55047				
ID	ABU55047	standard; Protein: 388	AA.	
XX	ABU55047;			
XX	18-MAR-2003 (first entry)			
DT	Human novel polypeptide #134.			
XX				
DE				
XX	Human; neural disorder; immune system disorder; renal disorder;			
KM	muscular disorder; respiratory disease; reproductive disorder;			
KM	gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;			
KM	hyperproliferative disorder; inflammatory disease; allergic reaction;			
KM	blood related disorder; cancer; immunosuppressive; antiinflammatory;			
KM	cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;			
KM	haemostatic; antiarteriosclerotic.			
XX				
OS	Homo sapiens.			
XX				
PN	US2002132753-A1.			
PD	19-SEP-2002.			
XX				

Db 309 ENIDVNEELPARRRRNRREDGKTEFVAQTVDPKRRRLQLDLGEYEVAKXMECEPISKKR 368
Qy 332 ATWETILDGKRLPFEPFESOGPTLOFTLRMTGETNDKSTADIAPIATRNSESJHOEKP 391
Db 369 ATWETILDGKRLPFEPFESOGPTLOFTLRMTGETNDKSTADIAPIATRNSESJHOEKP 428
Qy 392 GSVKPTQTIAVKESLTTDLQTRKEKDPNENRQKRLFTYQPLYNNTROQTEARDLHCP 451
Db 429 GSVKPTQTIAVKESLTTDLQTRKEKDPNENRQKRLFTYQPLYNNTROQTEARDLHCP 488
Qy 452 WCLTNCRLVSLNHLKLCRFRFTFNYYHPRGARIDVSIKCTDGSYAGNPDIHROP 511
Db 489 XXXXXXXXXXXXXXXXXXXXSRFTFNYYHPRGARIDVSIKCTDGSYAGNPDIHROP 548
Qy 512 FAFSRNGPVKRTPTITHILVCPKRTKASMESELESEDOEVOQRTYSSGHNRLFHSPTC 571
Db 549 FAFSRNGPVKRTPTITHILVCPKRTKASMESELESEDOEVOQRTYSSGHNRLFHSPTC 608
Qy 572 LPLRPQEMEVDSEDEKDEWLRKTTQIEEFSDVNEGEKVMKLMNLHVKHGFIADNQ 631
Db 609 LPLRPQEMEVDSEDEKDEWLRKTTQIEEFSDVNEGEKVMKLMNLHVKHGFIADNQ 668
Qy 632 MNHACMLEVNYGKIIKKNCIRNFMLHVMHDPNLISINSIDKATYKLEMOCKLEKG 691
Db 669 MNHACMLEVNYGKIIKKNCIRNFMLHVMHDPNLISINSIDKATYKLEMOCKLEKG 728
Qy 692 ESASPAHEITEENGTFANGFSEINSKKALETDSVSGVSKOSKKKL 739
Db 729 ESASPAHEITEENGTFANGFSEINSKKALETDSVSGVSKOSKKKL 776

RESULT 4
AAU15978
ID AAU15978 standard; Protein: 388 AA.
xx
AC AAU15978;
xx
DT 07-NOV-2001 (first entry)
xx
DE Human novel secreted protein, Seq ID 931.
xx
KW Human; immunosuppressive; antiarthritic; antirheumatic;
KW cytosolic; cardiant; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW vulnery; secreted protein; rheumatoid arthritis;
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW cerebrovascular disorder; cerebral ischemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW corneal infection; wound healing; epithelial cell proliferation;
KW skin ageing; food additive; preservative; antiproliferative.
xx
OS Homo sapiens.
xx
PN WO200153322-A2.
xx
PD 02-AUG-2001.
xx
PF 17-JAN-2001; 2001WO-US01341.
xx
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 06-SEP-2000; 2000US-0230439.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231245.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235835.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241121.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.

PT with altered vernalization response, flowering time, leaf size and/or
 PT shape or shade avoidance response for maximized reproductive success
 XX
 PS Disclosure; Page 79; 105pp; English.
 XX
 CC Isolated nucleic acid sequences obtained from the VRN2 locus of a
 CC plant encode polypeptides which are capable of affecting one or more
 CC vernalization responses such as, flowering time, leaf size and/or
 CC shape or the shade avoidance response of a plant into which the
 CC nucleic acid is introduced. Introducing such sequences into plants
 CC to alter these characteristics maximises the reproductive success of
 CC the plant. This protein is encoded an isolated human sequence which
 CC has homology to the Arabidopsis thaliana VRN2 gene over a short region
 CC near the N-terminus.
 CC
 SQ Sequence 803 AA;
 Query Match 100.0%; Score 3885; DB 21; Length 803;
 Best Local Similarity 100.0%; Pred. No. 7.8e-303;
 Matches 739; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAPQKGGGGGGGGGSGGSGGSGGSAVAATAAGKSGGSGGSGGSSASSSSA 60
 DB 65 MAPQKGGGGGGGGGSGGSGGSGGSAVAATAAGKSGGSGGSGGSSASSSSA 124
 QY 61 AAAAGAAVLPVKRKRMEHVAADHELFLQAEKPTQIYRFLTRNLIAPIFLHRTLYMSH 120
 DB 125 AAAAGAAVLPVKRKRMEHVAADHELFLQAEKPTQIYRFLTRNLIAPIFLHRTLYMSH 184
 QY 121 RNSFTNTRKRTFKYDDMLSKYERKMGEOESHLSAHQLPTFGFHNDKSPSENEON 180
 DB 185 RNSFTNTRKRTFKYDDMLSKYERKMGEOESHLSAHQLPTFGFHNDKSPSENEON 244
 QY 181 SVTLEVLIVKCHKKRKRKDVSCPIRQVPTGKKQVFLPIDLNOTKGNPSPSLAVSNEEPS 240
 DB 245 SVTLEVLIVKCHKKRKRKDVSCPIRQVPTGKKQVFLPIDLNOTKGNPSPSLAVSNEEPS 304
 QY 241 NSHWKYSYLLFRTVRGRREFGNGINGETNENIDVNEELPARRKRNREDGEKTVAMQT 300
 DB 305 NSHWKYSYLLFRTVRGRREFGNGINGETNENIDVNEELPARRKRNREDGEKTVAMQT 364
 QY 301 VEDKRRRLQLLDGEYEVANOMECPISKRRATWETLLDGKRLRPPEFTFSGPTLOFTLR 360
 DB 365 VEDKRRRLQLLDGEYEVANOMECPISKRRATWETLLDGKRLRPPEFTFSGPTLOFTLR 424
 QY 361 WTGETNDKSTAPIAKPLATRSESLHQBKPGSVKPTOTIAVKESTLTDLOTREKDTPN 420
 DB 425 WTGETNDKSTAPIAKPLATRSESLHQBKPGSVKPTOTIAVKESTLTDLOTREKDTPN 484
 QY 421 ENROKLRIFYOFLYNNNTROOTEARDLHCPWCTLNCRLYSLLKHLKCHSRPIFYVY 480
 DB 485 ENROKLRIFYOFLYNNNTROOTEARDLHCPWCTLNCRLYSLLKHLKCHSRPIFYVY 544
 QY 481 HPRGARLDVSTNECYDSYAGNPDHROGFARSRNGPKVKTITITILVCRPKRTASM 540
 DB 545 HPRGARLDVSTNECYDSYAGNPDHROGFARSRNGPKVKTITITILVCRPKRTASM 604
 QY 541 SEPLESEDEGEYOQRYSNGHNLRYFHSDDLPLRPOMEVDSDEKDEPDLARKTITOI 600
 DB 605 SEPLESEDEGEYOQRYSNGHNLRYFHSDDLPLRPOMEVDSDEKDEPDLARKTITOI 664
 QY 601 EEFSDVNEGEKEVKKLNHLVHKHGTADNOMNHACMLFVENYQKTIKKNLKCNFMHL 660
 DB 665 EEFSDVNEGEKEVKKLNHLVHKHGTADNOMNHACMLFVENYQKTIKKNLKCNFMHL 724
 QY 661 VSMHDFNLISIMSDKAVTKLRMOOKLGESESAPNEETTEQONTANGFSINSK 720
 DB 725 VSMHDFNLISIMSDKAVTKLRMOOKLGESESAPNEETTEQONTANGFSINSK 784
 QY 721 ALETDSVSGVSKSKOKL 739
 DB 785 ALETDSVSGVSKSKOKL 803

RESULT 2
 ABG30842
 ID ABG30842 standard; Protein: 739 AA.
 XX
 AC ABG30842;
 XX
 DT 21-OCT-2002 (first entry)
 XX
 DE Human joined with JAZF1 (JJA21) protein.
 XX
 KW Human: JAZF1; juxtaposed with another zinc finger: JJA21; JAZF1/JJA21;
 KW joined with JAZF1; proliferation; endometrial stroma tumour; immunogen;
 KW antigen; antibody; fertility; pregnancy; gene therapy; vaccine;
 KW chromosome 17.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Region 92..93
 FT /note= "Breakpoint for production of fusion protein
 FT JAZF1/JJA21"
 FT Misc-difference 448..471
 FT /note= "Encoded by CTGCATTGCCCTGTGTACTGTGACTGCGCA-
 FT ACTTATAGTTTACTCAAGCATCTTAACTGCGCAT.
 FT C2H2 zinc finger domain"
 FT Region 521..538
 FT /note= "Bipartite nuclear localisation signal"
 XX
 FN W0200193805-A2.
 XX
 PD 13-DEC-2001.
 XX
 PE 04-JUN-2001; 2001WO-0517936.
 XX
 PR 02-JUN-2000; 2000US-209093P.
 XX
 PA (BGM) BRIGHAM & WOMENS HOSPITAL INC.
 XX
 PI Koontz J, Sklar J;
 XX
 DR WPI: 2002-575047/61.
 DR N-PSDB: ABK89162.
 XX
 PT Novel JAZF1, JJA21 or JAZF1/JJA21 polypeptides useful as immunogens or
 PT antigens to raise or test anti-JAZF1, JJA21 or JAZF1/JJA21 antibodies
 PT
 PS Claim 5; Fig 2; 76pp; English.
 PS
 CC The present invention relates to a new JAZF1 (juxtaposed with another
 CC zinc finger), JJA21 (joined with JAZF1) or JAZF1/JJA21 polypeptide.
 CC The methods of the invention can be used to identify a compound which
 CC controls proliferation of endometrial stroma, by expressing JJA21 in the
 CC presence of the compound, and determining whether the compound affects
 CC expression of JJA21. JAZF1, JJA21 or JAZF1/JJA21 polypeptides are useful
 CC as immunogens or antigens to raise or test anti-JAZF1, JJA21 or
 CC JAZF1/JJA21 antibodies. The invention can be used as bait proteins in a
 CC two hybrid assay or three hybrid assay to identify other proteins which
 CC bind or interact with JAZF1/JJA21-binding proteins. JAZF1, JJA21 or
 CC JAZF1/JJA21 molecules are useful for identifying the origin of tumour
 CC and as tumour marker protein to verify that a stromal tumour is from
 CC endometrium. The antibody is useful for promoting or decreasing
 CC fertility or pregnancy, and also for treating endometrial stroma
 CC tumours. The present amino acid sequence represents the human JJA21
 CC protein of the invention. This sequence is encoded by the human JJA21
 CC gene located on chromosome 17.
 CC
 SQ Sequence 739 AA;
 Query Match 95.5%; Score 3709; DB 23; Length 739;
 Best Local Similarity 96.8%; Pred. No. 9.4e-289;
 Matches 715; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 2, 2003, 11:01:37 ; Search time 91.058 Seconds
(without alignments)
1288.178 Million cell updates/sec

Title: US-09-874-162a-5
Perfect score: 3885
Sequence: 1 MAPKRGCGGGGGGSGPSAGS.....KALETDSVGVSKSKKOKL 739

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 100%
Listing first 45 summaries

Database :

1: A.Geneseq.19Jun03.*
2: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1980.DAT.*
3: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1981.DAT.*
4: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1982.DAT.*
5: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1983.DAT.*
6: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1984.DAT.*
7: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1985.DAT.*
8: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1986.DAT.*
9: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1987.DAT.*
10: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1988.DAT.*
11: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1989.DAT.*
12: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1990.DAT.*
13: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1991.DAT.*
14: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1992.DAT.*
15: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1993.DAT.*
16: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1994.DAT.*
17: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1995.DAT.*
18: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1996.DAT.*
19: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1997.DAT.*
20: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1998.DAT.*
21: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1999.DAT.*
22: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA2000.DAT.*
23: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA2001.DAT.*
24: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA2002.DAT.*
25: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3885	100.0	803	21	AB000066 KIAA0160 polypepti
2	3709	95.5	739	23	ABG30842 Human joined with
3	3231	83.2	776	22	ABG30843 Human JAF1/JJA1
4	2043	52.6	388	22	AAU15978 Human novel secret
5	2043	52.6	388	24	ABU55047 Human novel polype
6	1497	38.5	289	22	AAU15958 Human novel secret
7	1497	38.5	289	24	ABU55027 Human novel polype
8	1453	37.4	292	22	AAU16416 Human novel secret
9	1453	37.4	292	24	ABU55485 Human novel polype

10	1177	30.3	955	22	ABB71635 Drosophila melanog
11	919	23.7	175	22	AAU16402 Human novel secret
12	919	23.7	175	24	ABU55471 Human novel polype
13	243.5	6.3	604	21	AAV53933 An Os-MPC1 protein
14	236	6.1	611	21	AAV53932 A MPC1 protein hav
15	230	5.9	498	21	AAV53932 Arabidopsis thaila
16	227.5	5.9	467	21	AAV53932 Arabidopsis thaila
17	221	5.7	445	21	AAV53932 Arabidopsis thaila
18	217	5.6	440	21	AAV53932 Arabidopsis thaila
19	190	4.9	692	24	VRN2 polypeptide.
20	190	4.9	692	24	VRN2 polypeptide.
21	162.5	4.2	813	23	AAV53932 Histone deacetylase
22	161	4.1	1390	23	AAV53932 F122 protein seque
23	159.5	4.1	342	22	AAV53932 Human ovarian can
24	159	4.1	339	21	AAV53932 Drosophila melanog
25	158	4.1	3969	15	AAV53932 Arabidopsis thaila
26	157.5	4.1	1315	22	AAV53932 Product of the GDN
27	154.5	4.0	2211	23	AAV53932 Drosophila melanog
28	153	3.9	549	22	AAV53932 Novel human diagno
29	153	3.9	549	22	AAV53932 Human translocatio
30	153	3.9	738	19	AAV53932 Drosophila melanog
31	147	3.8	831	16	AAV53932 Drosophila melanog
32	146.5	3.8	722	20	AAV53932 pMTSS1 MSP spider
33	146.5	3.8	1235	16	AAV53932 M. catarrhalis slr
34	146.5	3.8	1335	22	AAV53932 Potassium ion chan
35	146.5	3.8	1639	22	AAV53932 S cerevisiae TRK1.
36	146.5	3.8	1639	23	AAV53932 P. falciparum synt
37	146	3.8	907	22	AAV53932 Plasmodium falcipa
38	146	3.8	941	22	AAV53932 Intracellular traf
39	146	3.8	953	22	AAV53932 Intracellular traf
40	146	3.8	2230	24	AAV53932 Intracellular traf
41	145	3.7	1570	22	AAV53932 Protein differenti
42	145	3.7	1570	22	AAV53932 Drosophila melanog
43	144.5	3.7	615	20	AAV53932 Spider silk protei
44	144.5	3.7	953	22	AAV53932 Intracellular traf
45	144.5	3.7	1654	6	AAV50777 Sequence of the p1

ALIGNMENTS

RESULT 1
ID AAB00066 standard; Protein; 803 AA.
AC AAB00066;
XX 16-NOV-2000 (first entry)
DT 16-NOV-2000 (first entry)
XX KIAA0160 polypeptide.
DE Vernalization gene: VRN2; plant characteristic: flowering time;
KW leaf size; leaf shape; shade avoidance response; reproduction;
KW breeding; pollination; cultivation; human.
XX Homo sapiens.
OS WO200044918-A1.
XX 03-AUG-2000.
XX 28-JAN-2000; 2000WO-GB00248.
XX 28-JAN-1999; 99GB-0001927.
XX (PLAN-) PLANT BIOSCIENCE LTD.
XX Dean C, Gendall A;
XX MPI: 2000-499333/44.
XX N-PSDB; AAA47759.
XX Isolated vernalization gene VRN2 is used to produce transgenic plants